On parameterization of the inverse problem for estimating aquifer properties using tracer data

M. B. Kowalsky,[1] S. Finsterle,[1] K. H. Williams,[1] C. Murray,[2] M. Commer,[1] D. Newcomer,[2] A. Englert,[3] C. I. Steefel,[1] and S. S. Hubbard[1]

Received 26 July 2011; revised 27 February 2012; accepted 18 April 2012; published 28 June 2012.

[1] In developing a reliable approach for inferring hydrological properties through inverse modeling of tracer data, decisions made on how to parameterize heterogeneity (i.e., how to represent a heterogeneous distribution using a limited number of parameters that are amenable to estimation) are of paramount importance, as errors in the model structure are partly compensated for by estimating biased property values during the inversion. These biased estimates, while potentially providing an improved fit to the calibration data, may lead to wrong interpretations and conclusions and reduce the ability of the model to make reliable predictions. We consider the estimation of spatial variations in permeability and several other parameters through inverse modeling of tracer data, specifically synthetic and actual field data associated with the 2007 Winchester experiment from the Department of Energy Rifle site. Characterization is challenging due to the real-world complexities associated with field experiments in such a dynamic groundwater system. Our aim is to highlight and quantify the impact on inversion results of various decisions related to parameterization, such as the positioning of pilot points in a geostatistical parameterization; the handling of up-gradient regions; the inclusion of zonal information derived from geophysical data or core logs; extension from 2-D to 3-D; assumptions regarding the gradient direction, porosity, and the semivariogram function; and deteriorating experimental conditions. This work adds to the relatively limited number of studies that offer guidance on the use of pilot points in complex real-world experiments involving tracer data (as opposed to hydraulic head data).

Citation: Kowalsky, M. B., S. Finsterle, K. H. Williams, C. Murray, M. Commer, D. Newcomer, A. Englert, C. I. Steefel, and S. S. Hubbard (2012), On parameterization of the inverse problem for estimating aquifer properties using tracer data, *Water Resour. Res.*, *48*, W06535, doi:10.1029/2011WR011203.

1. Introduction

[2] Subsurface hydrological properties, such as permeability and porosity, can be inferred through inverse modeling of indirect measurements (e.g., hydraulic head and tracer concentrations) at discrete points and times, provided that such data are sufficiently sensitive to the properties of interest. However, the calibration of distributed groundwater models based on limited measurements is generally an underdetermined inverse problem [e.g., *McLaughlin and Townley*, 1996]. To overcome this limitation, simplifying assumptions regarding spatial variability are commonly made [e.g., *Moore and Doherty*, 2006]. For example, parameter zonation divides the model into a discrete number of zones, each with spatially uniform properties. The number and shape of zones, which may also contain parameterized property variations, may be iteratively determined during the calibration process with increasing granularity as supported by the calibration data [*Sun et al.*, 1998; *Tsai et al.*, 2003; *Berre et al.*, 2009], or a level set formulation that flexibly calibrates zone shapes during inversion of joint hydrogeophysical data may be used [*Cardiff and Kitanidis*, 2009]. Another option is to apply regularization, which enforces some form of spatial variability or smoothing in the property of interest, to make underdetermined inverse problems well posed [*Yeh*, 1986; *Carrera et al.*, 2005]. For example, a heterogeneous property can be cast as a spatially correlated random field if a statistical model of correlation (e.g., a semivariogram) can be inferred from site characterization data or concurrently estimated on the basis of the available calibration data [*Kitanidis*, 1995; *Hu*, 2000; *Caers*, 2003; *Finsterle and Kowalsky*, 2008]. While such simplifications may allow for a unique solution of the inverse problem, they result in a simplified picture of the subsurface that may or may not be adequate, depending on the application [*Moore and Doherty*, 2006]. Decisions regarding parameterization (i.e., how to represent a heterogeneous distribution with a limited number of parameters that are amenable to estimation) are of great importance for

[1]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA.
[2]Pacific Northwest National Laboratory, Richland, Washington, USA.
[3]Hydrogeology Department, Ruhr University Bochum, Bochum, Germany.

Corresponding author: M. B. Kowalsky, Earth Sciences Division, Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA. (mbkowalsky@lbl.gov).

This paper is not subject to U.S. copyright.
Published in 2012 by the American Geophysical Union

successful application of inverse modeling. In this study we aim to highlight the importance of proper spatial parameterization of subsurface heterogeneity, as errors in the model structure are partly compensated for by estimating biased property values during the inversion. These biased estimates—while potentially providing an improved fit to the calibration data—may lead to wrong interpretations and conclusions and reduce the ability of the model to make reliable predictions.

[3] Using at first a geostatistical parameterization, we treat the base 10 log permeability as a spatially correlated random field. To estimate its spatial distribution through inverse modeling of secondary measurements, we use an implementation of the popular pilot point method [*de Marsily*, 1978], of which many variations [e.g., *de Marsily et al.*, 1984; *Certes and de Marsily*, 1991; *Lavenue and Pickens*, 1992; *RamaRao et al.*, 1995; *Gómez-Hérnandez et al.*, 1997; *Doherty*, 2003; *Kowalsky et al.*, 2004, 2005; *Alcolea et al.*, 2006, 2008; *Finsterle and Kowalsky*, 2008] and related methods [*Rubin et al.*, 2010] have been proposed over the years, which have been reviewed by *Hendricks Franssen et al.* [2009]. Heterogeneous log permeability distributions are generated using sequential Gaussian simulation (SGSIM) [*Deutsch and Journel*, 1992], such that they reflect the spatial correlation specified by the semivariogram and such that they are conditioned to (i.e., a function of) so-called pilot point values, which are estimated in the inversion procedure. This geostatistical approach requires that the inversion procedure be repeated multiple times, each time with a different initial random field (based on a different seed number). The multiple inversion realizations result in a multitude of parameter distributions, which provide log permeability estimates at each location in the model with corresponding uncertainty [*RamaRao et al.*, 1995].

[4] While there is no set rule for determining the optimal positioning of pilot points for a given scenario, *Jung et al.* [2011] provide an excellent review of studies that have considered how to add pilot points (e.g., predefining them or adding them sequentially) and how best to select their locations, including empirically based approaches (e.g., random placement or a uniform density of pilot points) and sensitivity-based approaches that optimize pilot point placement on the basis of measurement locations (e.g., on the basis of the adjoint sensitivity technique of *Lavenue and Pickens* [1992] or the D optimality criterion proposed by *Jung et al.* [2011]). Motivated by the lack of studies offering guidelines on implementing pilot points in hydrogeological applications, *Doherty et al.* [2010] offer a variety of such guidelines based largely on the mathematical basis of the pilot point method, and they lay out some future related research directions. Among numerous recommendations, they cite the need for further synthetic studies to elucidate pilot point placement and other implementation details.

[5] The majority of pilot point applications rely on simplistic numerical experiments, and most limit their scope to hydraulic head measurements (and hydraulic conductivity measurements) rather than transient tracer measurements. In addition to the continued need for numerical experiments with pilot points in applications that are of practical relevance and contain a variety of data types, there is a need for more examples in which the methods are applied to field data from experiments with real-world complications and limitations, to help refine, improve and identify guidelines for successful inverse modeling.

[6] Heterogeneity can also be parameterized using geophysical data in a variety of ways, such as through tomographic constraints in the inversion of tracer data [*Linde et al.*, 2006]. Coupled hydrogeophysical approaches have combined traditional hydrological measurements with geophysical data, such as seismic data [*Hyndman et al.*, 1994; *Hyndman and Gorelick*, 1996], which are related to lithological zonation, or electrical resistivity data [e.g., *Pollock and Cirpka*, 2008; *Kowalsky et al.*, 2011; *Pollock and Cirpka*, 2010, 2012], which are sensitive to solute concentration and therefore provide secondary measurements that can be used to estimate hydrological properties.

[7] In general, a zonation parameterization is also of great value, as it is conducive to incorporating characterization data, such as from geophysical measurements, hydrological tests, or core data, into a model. Parameterization techniques can also be combined, such as in the zonation–kriging method of *Tsai* [2006] that integrates the conditional estimates of a kriged field within a geostatistical framework and of a zonal structure honoring a set of sampled data.

[8] *Dafflon et al.* [2011] evaluated several parameter estimation approaches for an application similar to the one considered here, a tracer experiment in an unconfined aquifer. Aside from gaining valuable insight into the variable-density flow phenomena in their experiment, resulting from the high concentration of saline tracer used, they evaluated the usefulness of various sources of geophysical and hydrological information. Their study demonstrated some of the challenges in dealing with real field experiments (e.g., their hydrological model could not properly fit the concentration breakthrough, which they speculate was due to inadequacies in the conceptual model or boundary conditions or in parameterization of heterogeneity).

[9] The current work is more directly motivated by a previous study (M. B. Kowalsky, S. Finsterle, A. Englert, K. H. Williams, C. Steefel, and S. S. Hubbard, Inversion of time-lapse tracer data for estimating changes in field-scale flow properties during biostimulation, submitted to *Journal of Hydrology*, 2012), which analyzed time-lapse tracer data collected in two consecutive biostimulation field experiments that were conducted in a flow cell at a uranium-contaminated aquifer at Rifle, Colorado, in 2002 and 2003. They performed hydrological inverse modeling of the tracer data, using a geostatistical parameterization, to estimate the heterogeneous log permeability distribution for each year. With a goal of identifying subtle changes in flow properties, such as those expected to occur during biostimulation, they concluded there was insufficient information in the tracer data of that particular experiment to accurately infer changes in permeability of less than half an order of magnitude from one year to the next. They hypothesized that the coarse well spacing of the experiment, relative to the length scale of heterogeneity, contributed to nonuniqueness in the inverse problem. The study also pointed to the need for a better understanding of how potential errors in the model parameterization could affect the solution of such inverse problems and how additional site characterization data might be included to reduce uncertainty in parameter estimates.

[10] This study is based on a subsequent field-scale tracer experiment conducted at the Rifle site. Described in detail by *Williams et al.* [2011], the experiment took place in 2007 within a flow cell having a closer well spacing than that used for the 2002–2003 study (Kowalsky et al., submitted, 2012). After providing some details about the site and the experiment in section 2, we describe the hydrological inverse modeling approach in section 3, including details of the hydrological model, parameterization techniques, and the inverse modeling procedure itself. Then we use a synthetic example in section 4 to examine how decisions made regarding parameterization impact solution of the inverse problem, and we examine issues that can arise when implementing zonation information, known to varying degrees of completeness and accuracy, such as from core logs or geophysical data, in the inversion procedure. The inverse modeling approaches are applied to actual field data in section 5, first using a 2-D model with a geostatistical parameterization and testing, among other things, the previous assumption of uniform porosity and gradient direction. The model is then extended to 3-D by employing a geostatistical parameterization together with a zonation parameterization that incorporates facies information derived from geologic well log descriptions, while accounting for uncertainty in the facies geometry. Comparisons are then made between permeability values that are estimated and derived from slug test data, and between values of porosity estimated in the study and inferred from other sources.

[11] The importance of this work is exemplified by the fact that one of the main difficulties in building reactive transport models for complex field sites continues to stem from uncertainty in the basic heterogeneous hydrological properties, such as permeability and porosity. Thus testing, improving, and refining techniques for estimating such properties continues to be an essential research topic in hydrogeology. Furthermore, there is a lack of studies highlighting the impact of decisions related to parameterization and quantifying how they affect inverse modeling results. We intend for this work to add to the relatively limited number of synthetic and field applications offering some guidance for the use of pilot points in complex real-world experiments involving tracer data (as opposed to hydraulic head data). The need is apparent for ongoing synthetic examples for testing the approach as new applications arise, for quantifying the impact of certain modeling assumptions, and for application to field data from complex real-world field experiments.

2. Description of Site and Experiment

[12] The experiment we consider was performed at the Department of Energy (DOE) Integrated Field Research Challenge Site (IFRC) at Rifle, Colorado, in a shallow unconfined aquifer contaminated from uranium mill tailings. The aquifer is 2.5 to 3 m thick, and is located on the floodplain of the Colorado river in alluvium situated above an impermeable bedrock formation (Wasatch) and below a clay fill layer that was put in place following removal of contaminated soil from the site. The IFRC site has been the subject of a number of experiments investigating uranium remediation through acetate biostimulation, a process involving the injection of an electron donor (acetate) to facilitate the microbial transformation of aqueous uranium, U(VI), to an insoluble form, U(IV) [e.g., *Anderson et al.*, 2003; *Yabusaki et al.*, 2007; *Englert et al.*, 2009; *Fang et al.*, 2009]. One focus of the IFRC experiments is to investigate whether the efficacy of biostimulation is impacted by subsurface heterogeneity, and whether biostimulation itself causes changes in flow properties that further influence its efficacy [*Li et al.*, 2010; Kowalsky et al., submitted, 2012]. To this end, nonreactive (i.e., conservative) tracers are typically injected into the groundwater to allow for characterization of the transport processes occurring in the experiments.

[13] Here we consider the "Winchester" experiment of 2007 [*Williams et al.*, 2011]. Site groundwater was amended with sodium bromide and acetate and mixed in a holding tank prior to injection within ten wells, screened over the entire aquifer and oriented approximately perpendicular to the predominant direction of groundwater flow. The resulting concentrations of various geochemical species were measured as a function of time in 12 downgradient monitoring wells that were also fully screened (Figure 1). Since the primary goal of the current study is to characterize heterogeneity of the hydrological properties, we focus on the analysis of the conservative tracer bromide. Figure 2 shows the approximate bromide concentration in the holding tank, along with the water and bromide injection rates, calculated on the basis of the concentration in the tank and the measured time-varying flow rate from the tank to the wells. Note that while the water table is known to fluctuate seasonally, it was nearly constant during the experiment (i.e., the saturated water thickness only varied by around 1%), allowing the aquifer to be modeled with a constant thickness of 2.5 m. These injection functions are used as input to the hydrological model for the synthetic example (section 4) and for application of the approach to the field data (section 5). The synthetic concentration data and the field data will be presented and discussed in the corresponding sections.

[14] Recent studies have analyzed data collected from the Winchester experiment within the context of reactive

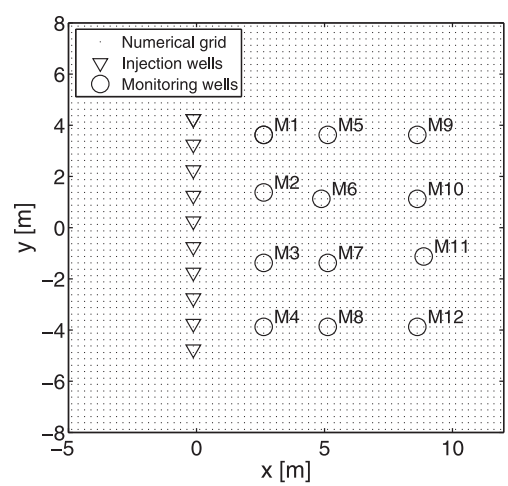

Figure 1. Numerical grid with locations of injection and monitoring wells for the 2007 "Winchester" experiment. The groundwater flow direction is approximately from left to right.

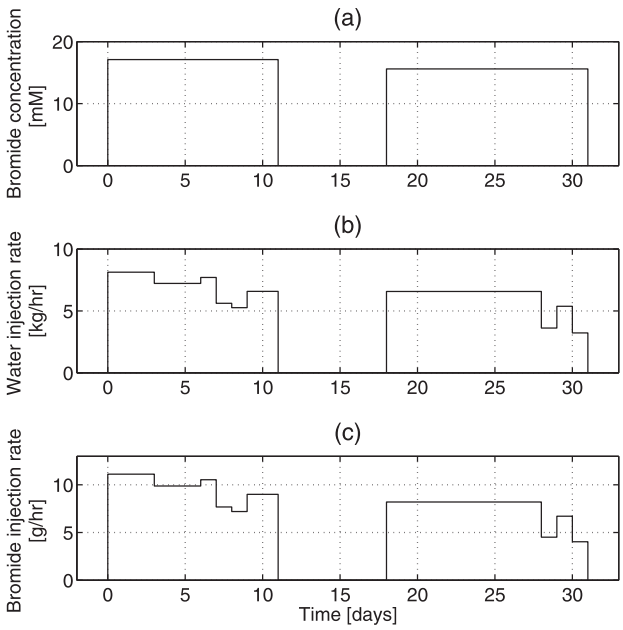

Figure 2. (a) Approximate bromide concentration in tank and rates at which (b) water and (c) bromide were pumped into the injection wells during the experiment.

transport modeling. *Fang et al.* [2009] developed a reaction network for the abiotic and microbially mediated reactions occurring during the biostimulation experiments, and performed biogeochemical reactive transport simulations. They calibrated the reaction network using a 1-D model with data from an earlier experiment, and tested its performance using later data sets, including those from the Winchester experiment. *Li et al.* [2011] performed 2-D reactive transport modeling, using permeability distributions obtained as part of the current study, to assess the relationship between physical and geochemical heterogeneity and bioremediation efficacy. *Yabusaki et al.* [2011] performed reactive transport modeling using 3-D lithofacies-based models of heterogeneity and found good agreement between the measured and simulated geochemical species for a second experiment conducted in 2008 in the same experimental field plot.

[15] In order to understand and accurately model the complex biogeochemical reactions occurring during biostimulation, such as at the Rifle site, it is crucial to be able to estimate heterogeneity in subsurface properties. Heterogeneity is expected to play a critical role in determining the efficacy of biostimulation. Therefore, methods for estimating heterogeneous subsurface properties on the basis of limited characterization data, such as tracer concentrations, core data, and geophysical data, are in high demand.

3. Approach

[16] In order to focus on a challenging setting that reflects the real world complexities associated with dynamic field experiments in shallow alluvial aquifers, we consider an inverse problem based on the Winchester experiment from the Rifle site described in section 2. In particular, we evaluate the use of inverse modeling of the bromide tracer data to estimate spatial variations in permeability. In most cases we assume that the porosity is constant and unknown, and we include its estimation in the inversion procedure. Since spatial variations in porosity can also have a significant impact on transport in some cases and therefore affect attempts at estimating permeability [e.g., *Hu et al.*, 2009; *Dafflon et al.*, 2011], we also allow for variable porosity in section 5.1.

[17] The inverse modeling involves the use of a forward model that simulates the tracer injection and the corresponding concentration measurements; model parameterization that defines how heterogeneity is represented in the model; and an inverse modeling procedure that allows for estimation of the unknown parameters using an optimization algorithm that minimizes the difference between the simulated and measured observations. Some details of the different components are given next.

3.1. Hydrological Model

[18] The experiment is modeled using the flow and transport simulator TOUGH2 [*Pruess et al.*, 1999]. While TOUGH2 can accommodate nonisothermal multiphase systems, the implementation used here is for the (2-D or 3-D) isothermal simulation of two mass components (water and bromide) in the aqueous phase. We neglect density-dependent flow due to the relatively low bromide concentrations used in the experiment. This is in contrast to the experiment of *Dafflon et al.* [2011] in which density-dependent flow effects were present because of the high concentration and high injection rate of the injected tracer (5 times higher and 5 orders of magnitude higher, respectively, than for the Winchester experiment). Hydrodynamic dispersion is neglected at present, under the assumption that pore level causes of dispersion are considered less significant, especially given model uncertainty, than the dispersion caused by heterogeneity, which is explicitly handled in the inversion. The 2-D model used in sections 4 and 5.1 is a depth-averaged model with a constant thickness of 2.5 m, and it covers a horizontal area of 17 m by 16 m, with constant grid spacing equal to 0.25 m, giving a total of 4352 grid blocks (see Figure 1). Recall that the injection and monitoring wells are screened over the entire aquifer, which helps justify simulating the experiment with a 2-D model. The 3-D model used in section 5.2 is similar except that it contains 12 grid blocks in the vertical direction (with 0.25 m spacing), giving a total of 52,224 grid blocks.

[19] We simulate injection of the bromide tracer by specifying the mass fluxes of water and bromide at each injection well on the basis of the concentration in the tank that supplies the tracer and on the measured flow rate from the tank to the wells (Figure 2), and we record the calculated concentrations at the monitoring wells at the specified times.

[20] While the groundwater flow direction and magnitude in general are known to fluctuate seasonally at the site, they were relatively constant during the Winchester experiment, allowing us to specify a fixed hydraulic head gradient of 0.004. For most cases, we assume the regional groundwater flow direction is perpendicular to the plane formed by the injection wells, and we fix the pressures at the up-gradient and down-gradient boundaries accordingly, and a zero-flux boundary condition is used for the boundaries parallel to the mean flow direction. We also consider a case in section 5.1 in which the gradient direction is varied, and

this requires that the boundary conditions be modified (as will be explained in section 5.1).

[21] It is worth noting that the local gradient in the vicinity of the injection wells can change during the experiments because of mounding of injectate within the injection wells; this effect is accounted for in the simulations.

3.2. Parameterization of the Inverse Problem

[22] The log permeability is described using a geostatistical parameterization (in sections 4.1–4.3 and 5.1), a zonation parameterization (in section 4.4), or a combination of the two (in sections 4.4 and 5.2). As mentioned, the porosity is considered spatially uniform for all cases except one in section 5.1. Note that any mention of log permeability throughout the paper refers to the base 10 logarithm of the permeability, which is given in m^2.

[23] In the geostatistical parameterization, the log permeability is treated as a spatially correlated field, which we assume can be described adequately using an exponential semivariogram model with a variance of 0.26 and an integral scale of 3.3 m on the basis of the analysis of slug test data from the site. (The integral scale is intentionally given an incorrect value of 6.6 m in one case in section 4.1.)

[24] We also explore zonation as an alternative parameterization of the inverse problem, wherein regions of the model with uniform material properties are defined on the basis of practical considerations (e.g., a uniform region is defined in a region of the model, such as up gradient of the injection wells) or on the basis of the geometry defined by characterization data, such as geophysical or core data, known to varying degrees of completeness and accuracy. The log permeability values for these regions are estimated in the inversion procedure. It may also be advantageous to employ the geostatistical parameterization in some regions of the model and the zonation parameterization elsewhere, such as when zonation information is only available for a subset of the entire model domain.

3.3. Inverse Modeling Procedure

[25] The code iTOUGH2 [*Finsterle*, 1999, 2004], which provides inverse modeling capabilities for TOUGH2, is used to estimate the unknown parameters in this study using a variation of the pilot point method described in section 1. The general procedure is as follows: (1) the parameter guesses are specified, (2) the initial pressure is obtained for the system by calculating the steady state distribution with the current parameter values and specified boundary conditions, (3) the initial tracer concentration is set uniformly to zero, (4) the simulation proceeds as bromide and groundwater are injected into the injection wells, (5) the simulated concentrations are recorded in each of the monitoring wells at the times when corresponding data are available, (6) the objective function is evaluated, (7) a nonlinear optimization algorithm, specifically the Levenberg-Marquardt algorithm [*Levenberg*, 1944; *Marquardt*, 1963], which aims to find the parameter values that minimize the objective function, is used to update the parameter guesses, and (8) the procedure is repeated, starting at step 1, until an estimate of the minimum value of the objective function is found. The objective function is a measure of the misfit between the measured and simulated data, formed by summing the squares of the weighted residuals, where a weighted residual is the difference between the simulated and measured data divided by the standard deviation of the measurement error. The concentration measurements represent the system state to be matched, and are weighted with a constant measurement error assumed to be given by a standard deviation of 0.31 mM. Note that in discussing the misfit between simulated and measured data, epistemic error is the more correct term (than measurement error), as it encompasses both inaccuracies in making the measurements and also, for example, model structural error [*Rubin*, 2003].

[26] Regarding the implementation of the pilot point method used, so-called prior data is ascribed to each pilot point with a value equal to the reference log permeability ($\log k_{ref}$) and a weight of one (corresponding to a log permeability standard deviation of 1.0), such that large deviations from this value are penalized in the objective function. Adding weight to the pilot points can result in a more stable inverse solution by reducing fluctuations of the pilot points that are insensitive to the observation data [*Alcolea et al.*, 2006]. Note that $\log k_{ref}$ is the mean value specified in the sequential Gaussian simulation procedure, though it does not always exactly equal the mean of the resulting distribution. It can be considered an unknown parameter in the inversion.

[27] To quantify how well model output reproduces measured data, we analyze statistical measures of the residuals (the difference between the measured and modeled output), as defined in Appendix A. In addition, we employ the measures of parameter uncertainty and estimation error that are given in Appendix B.

[28] It is worth describing the computational efforts required for this study. Many of the inversions were run on a Linux computer with an Intel(R) Xeon(R) Processor X5550 (clock speed of 2.67 GHz, 4 CPU cores, and 8.2 MB cache). For one typical case in which 51 parameters were estimated while running on a single processor, the average times for a forward run and an inversion realization were 1 s and 87 min, respectively. Since two series of inversion realizations were running in parallel on two processors, 34 inversion realizations could be completed in 28.5 h. In another case in which 82 parameters were estimated, the average times for a forward run and an inversion realization were 1.2 s and 102 min, respectively. In this case, since three series of inversions were running in parallel on three processors, 34 inversion realizations could be completed in 25.5 h. A Linux cluster was used in some cases, allowing the inversions to be further parallelized within iTOUGH2, such that a large number of processors equal to the total number of unknown parameters plus one could run in parallel, speeding up inversion realizations substantially.

4. Synthetic Example

[29] In this section we consider a synthetic example based on the Winchester field experiment which represents a realistic experiment in a dynamic alluvial aquifer with real-world complexities and limitations. In sections 4.1 and 4.2, we focus on several inversion cases to highlight and quantify the impact of key decisions for parameterizing the inverse problem on parameter estimates and on the ability

of the model to reproduce the synthetic data. We also examine the performance of the pilot point approach for this particular application in the face of deteriorating experimental conditions (decreased sampling frequency and increasing measurement error) in section 4.3.

[30] In anticipation of requirements to extend such models from 2-D to 3-D with the consequent requirement to estimate 3-D heterogeneous property distributions, nonunique parameter estimates are expected (especially when tracer data represent depth-averaged measurements), making it desirable to consider alternative parameterizations (i.e., in addition to the geostatistical parameterization with the pilot point approach) that can directly integrate zonal information derived from geophysical data or core logs. Therefore, we take advantage of the synthetic example to examine some decisions made in implementing zonal information into a model instead of (or in addition to) parameterization with the pilot point method. In particular, we examine the potential for using a zonation parameterization in which the geometry of the main features of heterogeneity may be perfectly known to varying degrees, known inaccurately, or known only over a subset of the model domain (section 4.4).

[31] The synthetic concentration data were generated using the forward model described in section 3.1 with a 2-D permeability distribution (Figure 3) that, incidentally, was obtained from a single inversion realization of an early attempt at inversion of the actual field data. Using such a distribution as the "true" model for the synthetic example has the benefit of providing synthetic data that are similar to those of an actual field experiment. The porosity of the synthetic model is 0.121.

[32] Figure 4 shows simulated tracer concentrations and corresponding changes in pressure due to the tracer injection at several times. As is evident in Figures 4a and 4b, the injection rate is not high enough relative to the groundwater velocity to cause up-gradient migration of the tracer (this is consistent with the Winchester field experiment, where no tracer was detected in wells located slightly up gradient of the injection wells). Relatively minor increases in pressure are observed (Figures 4d and 4e), with increases highest near the injection wells and dropping off with distance from the injection wells. Relative to preinjection conditions, pressures increase by around 40% of the "maximum pressure difference" in the model (the difference between the pressure at the left and right boundaries), with the largest increase occurring in a low-permeability zone near x = 0 m and y = 2.5 m. By 50 days, when no tracer injection is occurring, the pressure has returned to preinjection conditions (Figure 4f), though the tail end of the bromide tracer is still evident (Figure 4c). We expect that the location of the model boundaries will not influence the simulation results substantially. Implementing the approach of *Lehikoinen et al.* [2007] would be an excellent way to account for errors due to the truncation of the computational domain, but this is left for future research. At present, we simply take advantage of the fact that the pressures and concentrations are known for the synthetic model. (For the application to field data in section 5, we make the same assumption, but we vary the gradient direction in one case.)

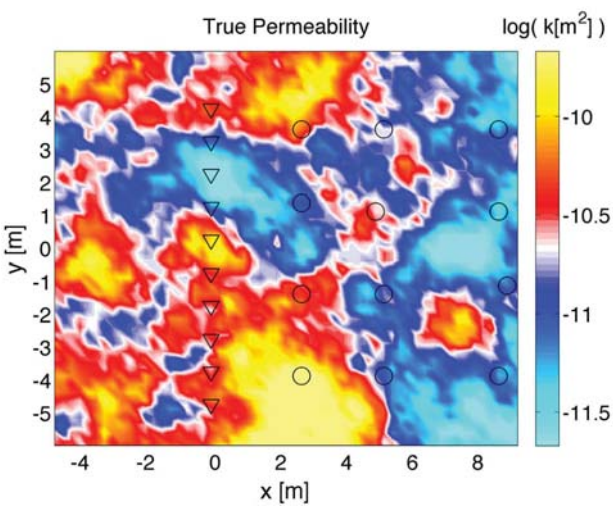

Figure 3. Permeability distribution used for the synthetic example in section 4. The injection and monitoring wells are shown with triangles and circles, respectively. Note that to improve visualization, only a subset of the entire modeling domain (see Figure 1) is shown here and in the remaining figures that show 2-D permeability distributions (Figures 6, 7, 8, 9, and 12).

[33] We consider two sampling scenarios: one with fine sampling (every two days between days 2 and 62 for a total of 31 times), referred to as "31t", and one with the comparatively coarse sampling of the actual field experiment (on days 2, 4, 6, 11, 15, 18, 20, 23, 26, 36, 41, 62 for a total of 12 times), referred to as "12t." For sampling scenario 31t, which is considered in sections 4.1 and 4.2, the synthetic data are free of measurement error (see Figure 5a). For sampling scenario 12t, which is considered in sections 4.3 and 4.4, three versions of synthetic data are used: one with no measurement error (Figure 5b), referred to as "N0"; one with a low level of measurement error (Figure 5c), referred to as "N1"; and one with a high level of measurement error (Figure 5d), referred to as "N2." Measurement error is introduced in the synthetic data by adding to it uncorrelated random noise with a mean of zero and a standard deviation of 0.13 and 0.31 mM for N1 and N2, respectively.

[34] Details for the different inversion cases of sections 4.1 and 4.2 are given in Table 1, while the cases of sections 4.3 and 4.4 are given in Tables 2 and 3.

4.1. Sensitivity to Pilot Point Spacing and Integral Scale of Semivariogram (With 31 Sampling Times and No Measurement Error)

[35] When using a pilot point parameterization, inversion results are potentially sensitive to choices made regarding the implementation of pilot points (e.g., how many of them are used, what regions they cover, and what their spacing is). The spacing of pilot points is typically set to around half the integral scale of the heterogeneity (as mentioned in section 3.2, a value of 3.3 m is used for the integral scale in this study), though optimal spacing may depend on details of the particular application. In addition, one must take into account the tradeoff between the desired resolution of the estimated distribution, and the potentially limited amount of complementary information contained in the data. The parameters of interest must have sufficient sensitivity to the concentration data to allow for their estimation. *Doherty*

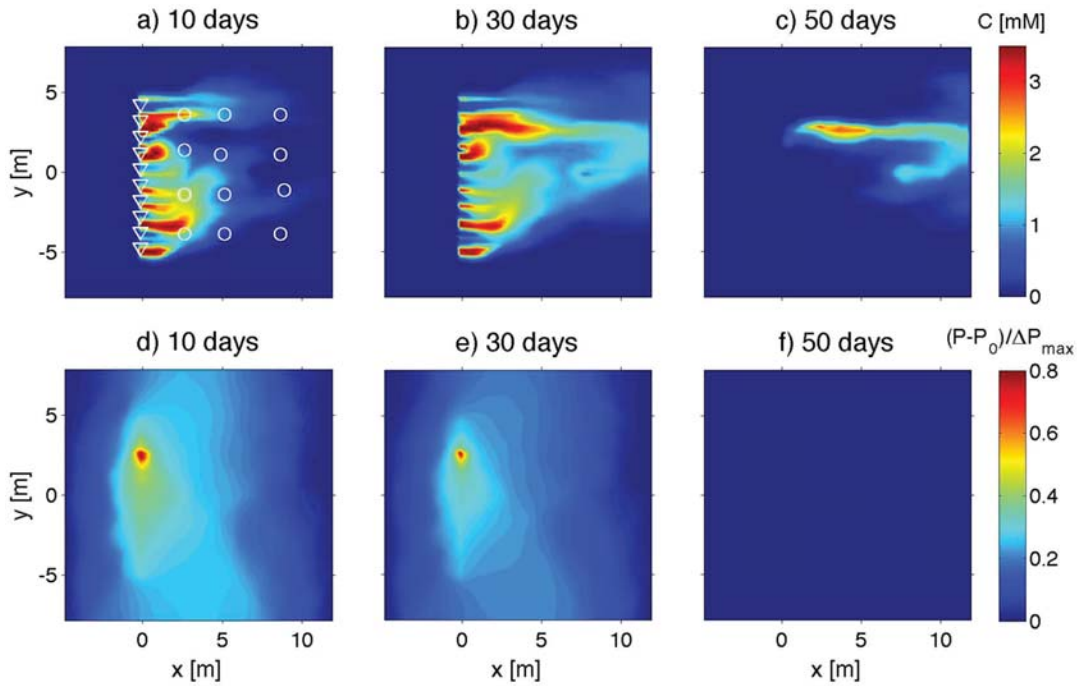

Figure 4. (a–c) Snapshot of tracer concentration and (d–f) corresponding normalized change in pressure due to injection at 10, 30, and 50 days, respectively, for the synthetic example considered in section 4. The normalized change in pressure is given by $(P - P_0)/\Delta P_{max}$, where P is the simulated pressure at any given location at the indicated time, P_0 is the initial pressure before injection at the same location, and ΔP_{max} is the maximum difference in pressure in the model (equal to the difference in pressure between the left and right boundaries). The injection and monitoring wells are shown in Figure 4a with triangles and circles, respectively.

et al. [2010] describe the goal of making a network that is dense enough for differences between the interpolated values between pilot points and the true parameter field to lie in the "calibration null-space," in other words, so that they are not informed by the observations.

[36] We test the inversion using four pilot point configurations that we refer to, for the sake of discussion, as having "coarse," "medium," "medium-shifted," and "fine" pilot point spacing (cases 31t-N0-C, 31t-N0-M, 31t-N0-M-shifted, and 31t-N0-F, respectively), as depicted in Figures 6a, 6c, 6e, and 6g. The coarse case (31t-N0-C) uses a total of 35 pilot points in five rows and seven columns with 2 m spacing (or 8 grid blocks) in both directions. The medium case (31t-N0-M) uses a total of 49 pilot points in seven rows and seven columns with the same spacing of columns (2 m or 8 grid blocks) but decreased spacing of rows (1.75 m or 7 grid blocks). The medium-shifted case uses the number of pilot points as in the medium case but every other column of pilot points shifted in the y direction by 0.75 m (or 3 grid blocks). The fine case (31t-N0-F) uses a total of 80 pilot points in eight rows and ten columns with spacing in both directions decreased further (1.5 m or 6 grid blocks). Each case in this section uses the finely sampled synthetic data (31t) with no measurement error (N0), representing ideal conditions.

[37] In addition, an inversion case (31t-N0-Uni) is conducted in which the permeability and porosity are homogeneous and their values are estimated in the inversion. This case serves as a baseline for the various performance measures reported in Table 1, and it provides us a means to assess the degree of heterogeneity in the model.

[38] The inversion results are summarized in Table 1. The homogeneous model does not allow for an adequate fit between the synthetic data and the simulated concentrations. For the heterogeneous cases, the fit improves substantially by decreasing the spacing of pilot points by only 25 cm (or 1 grid block). For example, when going from coarse to medium spacing, the error variance s_0^2 (equation (A1)), which describes the fit to the observations (i.e., the synthetic data in section 4), drops from 0.207 to 0.132 (Table 1). For reference, the corresponding value for the homogeneous model is 1.66. Keeping the same spacing of columns but vertically shifting every column slightly results in a value of s_0^2 (0.136) that is not improved relative to the case with medium spacing. Changing from medium to fine spacing results in s_0^2 decreasing to 0.126, which is only a minor improvement, especially considering that the standard deviation of s_0^2 is over 0.04. The synthetic data and simulated concentrations for the case with medium spacing (31t-N0-M) are shown in Figure 5a, and the simulated concentrations for the homogeneous reference case are shown as well.

[39] The overall appearance of the estimated log permeability (Figure 6) for each heterogeneous case is similar to

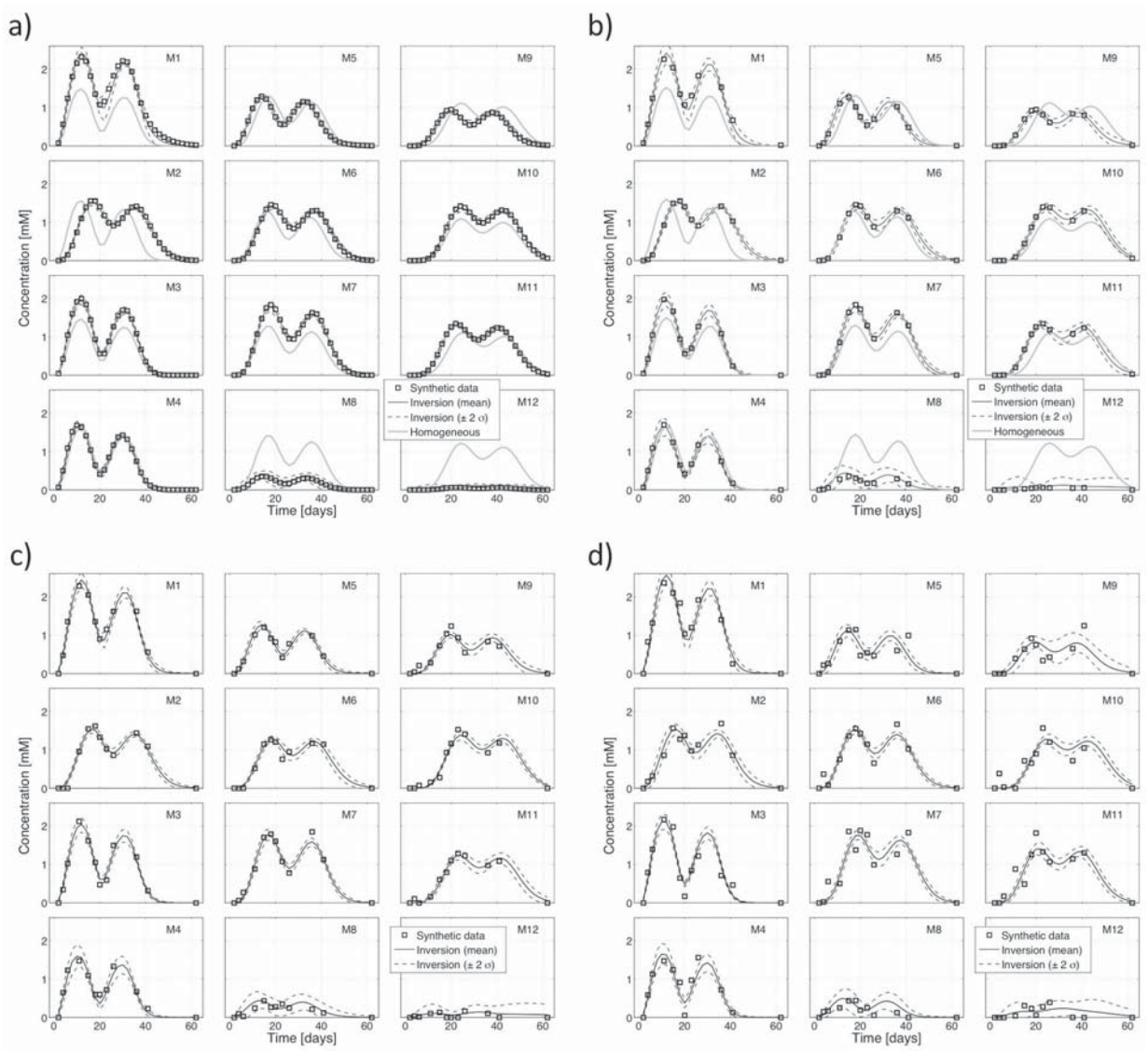

Figure 5. Examples of bromide concentrations for synthetic inversion cases that use the medium pilot point spacing: (a) 31 sampling times and no measurement error (case 31t-N0-C in section 4.1) and 12 sampling times with (b) no measurement error (case 12t-N0-M in section 4.3), (c) low measurement error (case 12t-N1-M in section 4.3), and (d) high measurement error (case 12t-N2-M in section 4.3). For monitoring wells M1 to M12, the synthetic data are shown with symbols, and the inversion results are shown with solid lines (mean of 30 realizations) and dashed lines (mean ± 2 standard deviations). For reference, the homogeneous cases 31t-N0-Uni and 12t-N0-Uni are shown with solid gray lines in Figures 5a and 5b, respectively.

that of the true model (Figure 3). As listed in Table 1, various performance measures are calculated for the region covered by the injection and monitoring wells (within the area given by $-0.5 < x < 9.5$ and $-4.5 < y < 4.5$). The average standard deviation of the estimated log permeability $\sigma_{\log k,ave}$ (equation (B3)) is lowest, equal to 0.411, for the case with fine pilot point spacing. However, the case with medium spacing (31t-N0-M) provides slightly more accurate estimates based on the log permeability estimate error values ε_1 and ε_2 (equations (B4) and (B5), respectively).

[40] It is possible that the specific locations of pilot points could explain why the medium spacing case provides slightly better results. We added cases for the medium and fine spacing where the pilot points were shifted by a constant value of 0.25 and 0.75 cm, respectively, in both the x and y directions. However, while the values varied slightly with the shifted cases, the results remained consistent with the results for medium spacing fitting slightly better than for the fine spacing. In addition, a case (31t-N0-M-Shifted) in which every other column of pilot points is

Table 1. Summary of 2-D Inversions for Synthetic Example

	Case							
	31t-N0-Uni	31t-N0-C	31t-N0-M	31t-N0-M-Shifted	31t-N0-F	31t-N0-M-Error	31t-N0-MH	31t-N0-MU
Section	4.1	4.1	4.1	4.1	4.1	4.1	4.2	4.2
Description	Uniform reference case	Coarse pilot point spacing	Medium pilot point spacing	Shifted pilot point locations	Fine pilot point spacing	Incorrect correlation length	Up-gradient region has no pilot points	Up-gradient region is uniform
Number of observation times	31	31	31	31	31	31	31	31
Number of bromide data m_b	372	372	372	372	372	372	372	372
Measurement error σ_{bromide}[a] (mmol)	0	0	0	0	0	0	0	0
Number of pilot points m_{pp}[b]	0	35	49	49	80	49	35	35
Number of unknowns n	2	37	51	51	82	51	37	38
$m_b + m_{pp} - n$ (equation (A1))	370	370	370	370	370	370	370	369
Parameter estimates								
Pilot points	N/A	Figures 6a and 6b	Figures 6c and 6d	Figures 6e and 6f	Figures 6f and 6g	Figures 6i and 6j	Figures 7a and 7b	Figures 7c and 7d
$\sigma_{\log k,ave}$ (equation (B3))	0.0	0.609	0.501	0.500	0.411	0.432	0.564	0.517
$\log k_{ref}$	−10.742 (±0.012)[c]	−10.642 (±0.048)	−10.643 (±0.043)	−10.626 (±0.040)	−10.626 (±0.018)	−10.643 (±0.033)	−10.648 (±0.060)	−10.660 (±0.053)
ϕ	0.15 (±0.005)	0.118 (±0.013)	0.121 (±0.013)	0.123 (±0.011)	0.120 (±0.008)	0.119 (±0.011)	0.116 (±0.011)	0.117 (±0.011)
Estimation error								
$\log k$ error ε_1 (equation (B4))	0.374	0.358	0.277	0.285	0.297	0.286	0.307	0.316
$\log k$ error ε_2 (equation (B5))	0.227	0.205	0.127	0.142	0.148	0.134	0.156	0.165
Residual analysis								
Error variance s_0^2 (equation (A1))	1.66	0.207 (±0.067)	0.132 (±0.049)	0.136 (±0.046)	0.126 (±0.044)	0.126 (±0.04)	0.178 (±0.071)	0.166 (±0.054)
NS (equation (A2))[d]	0.539	0.977 (±0.017)	0.992 (±0.006)	0.989 (±0.009)	0.994 (±0.007)	0.989 (±0.010)	0.981 (±0.014)	0.984 (±0.013)

[a]Standard deviation of the zero-mean measurement error that was added to the synthetic data in section 4.
[b]The pilot point parameters (log permeability values) are given prior data equal to $\log k_{ref}$, with a standard deviation of 1.0.
[c]Average value for all realizations followed in parentheses by the standard deviation (where applicable).
[d]Nash-Sutcliffe model efficiency.

Table 2. Summary of 2-D Inversions for Synthetic Example

	Case			
	12t-N0-Uni	12t-N0-M	12t-N1-M	12t-N2-M
Section	4.3	4.3	4.3	4.3
Description	Uniform reference case	Reduced observation times, no noise	Reduced observation times, low noise	Reduced observation times, high noise
Number of observation times	12	12	12	12
Number of bromide data m_b	144	144	144	144
Measurement error σ_{bromide}[a] (mmol)	0	0	0.13	0.31
Number of pilot points m_{pp}[b]	0	49	49	49
Number of unknowns n	2	51	51	51
$m_b + m_{pp} - n$ (equation (A1))	142	142	142	142
Parameter estimates				
Pilot points	N/A	Figures 8a and 8b	Figures 8c and 8d	Figures 8e and 8f
$\sigma_{\log k,ave}$ (equation (B3))	0.0	0.437	0.436	0.430
$\log k_{ref}$	−10.758 (±0.021)[c]	−10.636 (±0.031)	−10.638 (±0.037)	−10.640 (±0.026)
ϕ	0.155 (±0.008)	0.122 (±0.011)	0.122 (±0.013)	0.113 (±0.010)
Estimation error				
log k error ε_1 (equation (B4))	0.371	0.292	0.295	0.308
log k error ε_2 (equation (B5))	0.226	0.145	0.140	0.162
Residual analysis				
Error variance s_0^2 (equation (A1))	1.85	0.242 (±0.078)	0.307 (±0.077)	0.649 (±0.086)
NS (equation (A2))[d]	0.479	0.864 (±0.017)	0.899 (±0.012)	0.867 (±0.012)

[a]Standard deviation of the zero-mean measurement error that was added to the synthetic data in section 4.
[b]The pilot point parameters (log permeability values) are given prior data equal to log k_{ref}, with a standard deviation of 1.0.
[c]Average value for all realizations followed in parentheses by the standard deviation (where applicable).
[d]Nash-Sutcliffe model efficiency.

shifted by ±0.75 m in the y direction slightly worsened the fit to the data and the accuracy of the estimates, suggesting that regular spacing is preferred for this type of experiment.

[41] In all heterogeneous cases, the estimated porosity values are within one standard deviation of the true value (0.121), while the value for the homogeneous case is significantly inaccurate. Recall that log k_{ref} is the mean value specified in the sequential Gaussian simulation procedure, and it does not necessarily equal the mean of the resulting distribution. Accordingly, the estimated values of log k_{ref} for each case range between −10.642 and −10.626, which is closer to the actual mean of the true log permeability distribution (−10.677) than to the reference value used to generate the true distribution (−10.558).

[42] We assume that the integral scale of the semivariogram is known accurately for the inversion, but we were curious whether its value could be estimated as a parameter in the inversion and whether an inaccurately assumed value would negatively affect the parameter estimates. We found that there was insufficient sensitivity of the concentration data to allow for estimation of the integral scale in the inversion. Furthermore, when purposefully assigning an inaccurate value (6.6 m instead of the true value of 3.3 m), the resulting errors in the inversion results were minor (see case 31t-N0-M-Error in Table 1).

4.2. Sensitivity to Parameterization of Up-Gradient Region (With 31 Sampling Times and No Measurement Error)

[43] In section 4.1, we considered several cases with pilot points covering not only the region of the injection and monitoring wells but also the region up gradient of the injection wells (i.e., between the left boundary of the model and the injection wells, as seen in Figure 1). In this section we explore two alternatives for parameterizing the "up-gradient region" in such a way as to reduce the number of parameters to be estimated in the inversion. For this purpose we use a pilot point configuration with medium spacing, similar to case 31t-N0-M from section 4.1 except that two columns of pilot points are removed from the up-gradient region, reducing the number of pilot points from 49 to 35. The up-gradient region is handled in one of two ways. In the first case (case 31t-N0-MH), it is given unconditional heterogeneity (i.e., the log permeability is heterogeneous and contains the specified spatial correlation, on average, but it is not influenced by any pilot point values in that region). In the second case (case 31t-N0-MU), the log permeability in the up-gradient region is made uniform and its value is estimated in the inversion. Inversion proceeds as before, with both cases using the fine sampling scenario (31t) with no measurement error (N0), again representing ideal conditions.

[44] The inversion results are summarized in Table 1, and the log permeability estimates are shown in Figure 7. The fit to the observations is worsened, relative to the case from section 4.1 with medium spacing (case 31t-N0-M), with the error variance increasing by 35% and 26%, respectively, for the cases with the up-gradient region modeled using unconditional heterogeneity (case 31t-N0-MH; Figures 7a and 7b) and a uniform value (case 31t-N0-MU; Figures 7c and 7d). The average standard deviation $\sigma_{\log k,ave}$ increases by 13% and 3% for the respective cases. In addition, the error in log k increases for the first case, by 11% for ε_1 (equation (B4)) and 23% for ε_2 (equation (B5)), and for the second case by 14% for ε_1 and 30% for ε_2.

[45] While an inverse modeling practitioner might naively choose to handle an up-gradient region in a similar manner, perhaps with the goal of reducing the number of parameters or because it is well outside of the region of interest and not thought to be important, this example

Table 3. Summary of 2-D Inversions for Synthetic Example

	Case				
	12t-N0-3Z	12t-N0-5Z	12t-N0-5Z-Error	12t-N0-5Zlim	12t-N0-5ZlimPP
Section	4.4	4.4	4.4	4.4	4.4
Description	Three zones perfectly known	Five zones perfectly known	Five zones inaccurately known (errors up to 50 cm)	Five zones known for limited region; homogeneous outside of wells	Five zones known for limited region; pilot points outside of wells
Region modeled with five zones (or facies)	Entire model	Entire model	Entire model	-1 m $< x <$ 12 m, -5 m $< y <$ 5 m	-1 m $< x <$ 12 m, -5 m $< y <$ 5 m
Region modeled as uniform (homogeneous)	N/A	N/A	N/A	For $x < -1$ m, all y; for $x > -1$ m, $y < -5$ m, $y > 5$ m	N/A
Region modeled with pilot points	N/A	N/A	N/A	N/A	For $x < -1$ m, all y; for $x > -1$ m, $y < -5$ m, $y > 5$ m
Number of observation times	12	12	12	12	12
Number of bromide data m_b	144	144	144	144	144
Measurement error, σ_{bromide}[a] (mmol)	0	0	0	0	0
Number of pilot points m_{pp}[b]	0	0	0	0	24
Number of unknowns n	6	6	6	7	31
$m_b + m_{pp} - n$ (equation (A1))	138	138	138	137	137
Parameter estimates					
Pilot points	N/A	N/A	N/A	N/A	Figure 9f
log k_{ref} (for pilot points)	N/A	N/A	N/A	N/A	-10.57 (± 0.17)[d]
log $k_{upgradient}$	N/A	N/A	N/A	-11.06 (± 0.05)[c]	N/A
log k_1 (for facies 1)	-10.07 (± 0.03)[c]	-9.65 (± 0.03)[c]	-10.07 (± 0.05)[c]	-8.94 (± 0.36)[c]	-9.60 (± 0.01)[d]
log k_2 (for facies 2)	-10.51 (± 0.03)[c]	-10.22 (± 0.02)[c]	-9.39 (± 0.05)[c]	-10.24 (± 0.11)[c]	-10.22 (± 0.31)[d]
log k_3 (for facies 3)	-11.49 (± 0.03)[c]	-10.61 (± 0.01)[c]	-10.79 (± 0.03)[c]	-10.55 (± 0.09)[c]	-10.61 (± 0.11)[d]
log k_4 (for facies 4)	N/A	-11.22 (± 0.01)[c]	-10.98 (± 0.03)[c]	-10.70 (± 0.09)[c]	-11.17 (± 0.07)[d]
log k_5 (for facies 5)	N/A	-11.81 (± 0.03)[c]	-12.87 (± 0.08)[c]	-10.88 (± 0.22)[c]	-11.73 (± 0.05)[d]
ϕ	0.115 (± 0.003)[c]	0.121 (± 0.001)[c]	0.132 (± 0.004)[c]	0.133 (± 0.006)[c]	0.120 (± 0.19)[d]
Estimation error					
log k error, ε_1 (equation (B4))	0.187	0.172	0.214	0.334	0.164
log k error, ε_2 (equation (B5))	0.050	0.044	0.070	0.172	0.040
Residual analysis					
Error variance s_0^2 (equation (A1))	0.767	0.051	0.419	1.17	0.224 (± 0.083)[d]
NS (equation (A2))[e]	0.813	0.906	0.814	0.652	0.910 (± 0.012)[d]

[a]Standard deviation of the zero-mean measurement error that was added to the synthetic data in section 4.
[b]The pilot point parameters (log permeability values) are given prior data equal to log k_{ref}, with a standard deviation of 1.0.
[c]Estimated values followed in parentheses by the standard deviation.
[d]Estimated values given by the mean of 30 inversion realizations, followed in parentheses by the standard deviation.
[e]Nash-Sutcliffe model efficiency.

illustrates the high sensitivity of the results to the properties in this region. This example also points to the fact that a rigorous strategy for gauging sensitivity, such as using an adjoint state method [e.g., *Cirpka and Kitanidis*, 2001] could provide similar additional insights into the sensitivity of tracer measurements to various spatially distributed parameters in the hydrological model.

4.3. Sensitivity to Sampling Times and Measurement Error (With 12 Sampling Times and Variable Measurement Error)

[46] To examine the performance of the pilot point approach for this particular application in the face of less ideal experimental conditions, next we quantify the impact on the inversion results of decreasing the number of sampling times and adding measurement error to the concentration data. Specifically, we switch from the idealized sampling scenario of 31t to the limited scenario of 12t, which uses the same 12 sampling times as in the actual field experiment, and we consider three cases with differing amounts of measurement error: N0, N1, and N2 (cases 12t-N0-M, 12t-N1-M, and 12t-N2-M, respectively). The inversions for these cases use 49 pilot points with the medium spacing (i.e., the pilot point configuration used in case 31t-N0-M of section 4.1). An example (for case 12t-N1-M) of the estimated mean and standard deviation of the permeability distribution is shown in Figure 8. It is reassuring that the overall appearance of the log permeability remains similar to that of the true model. The inversion results for each case are summarized in Table 2.

[47] The main impact of reducing the number of survey times from 31 to 12, on the basis of comparison of cases 31t-N0-M and 12t-N0-M (see Table 2), is an increase in the error variance s_0^2 from 0.132 to 0.242, and also a slight increase in the log permeability error (ε_1 and ε_2 increase by 5% and 14%, respectively). Figures 5a and 5b allow for visual comparison of the fit between the synthetic data and the simulated concentrations.

[48] With the addition of measurement error, the fit between the synthetic data and the simulated concentrations worsens; the error variance s_0^2 increases from 0.242 to 0.307 and to 0.649 for cases 12t-N1-M and 12t-N2-M, respectively

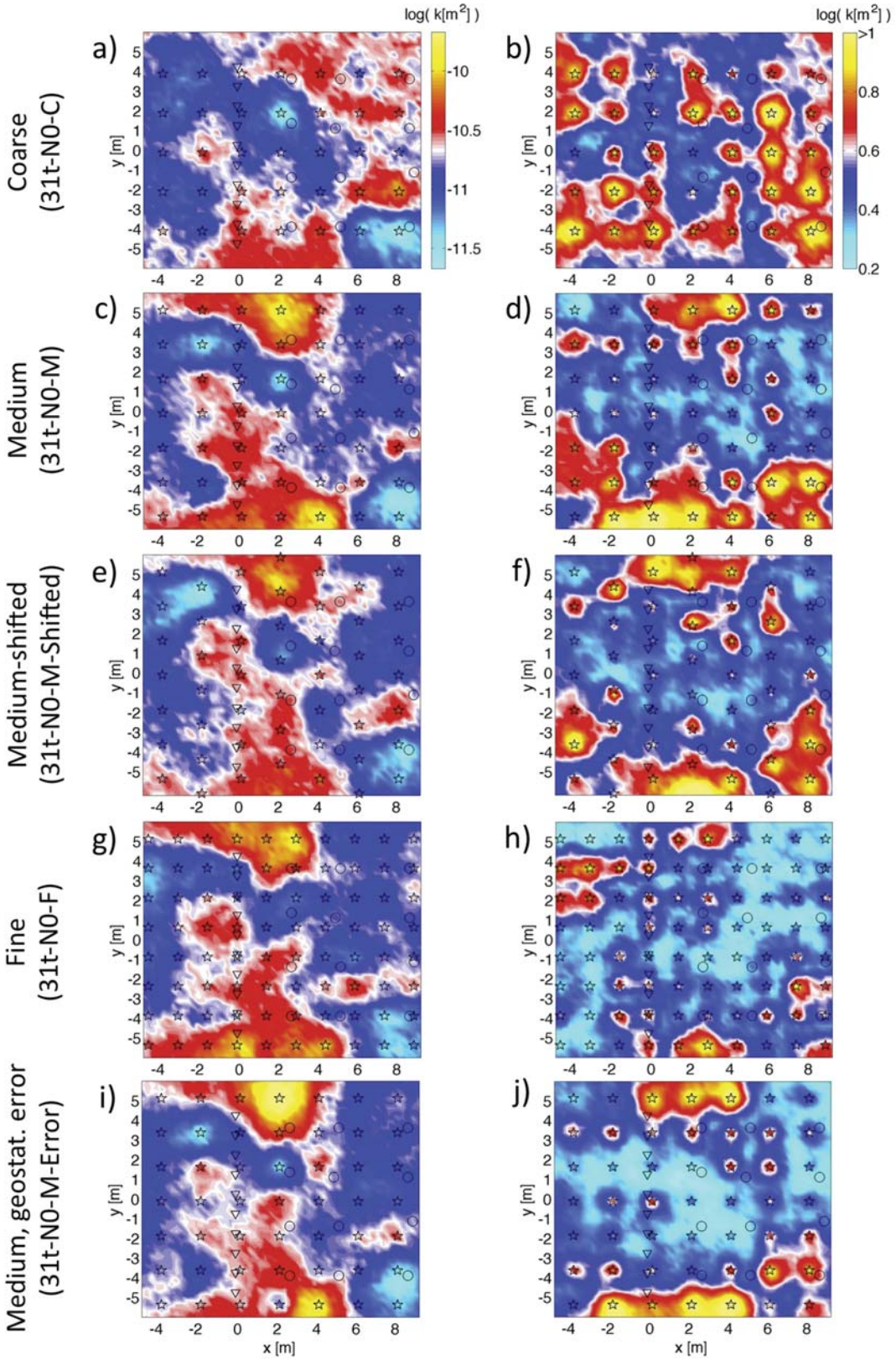
log(k[m^2])
a)
b)
Coarse
(31t-N0-C)
c)
d)
Medium
(31t-N0-M)
e)
f)
Medium-shifted
(31t-N0-M-Shifted)
g)
h)
Fine
(31t-N0-F)
i)
j)
Medium, geostat. error
(31t-N0-M-Error)
y [m]
x [m]
-10
-10.5
-11
-11.5
>1
0.8
0.6
0.4
0.2

Figure 6

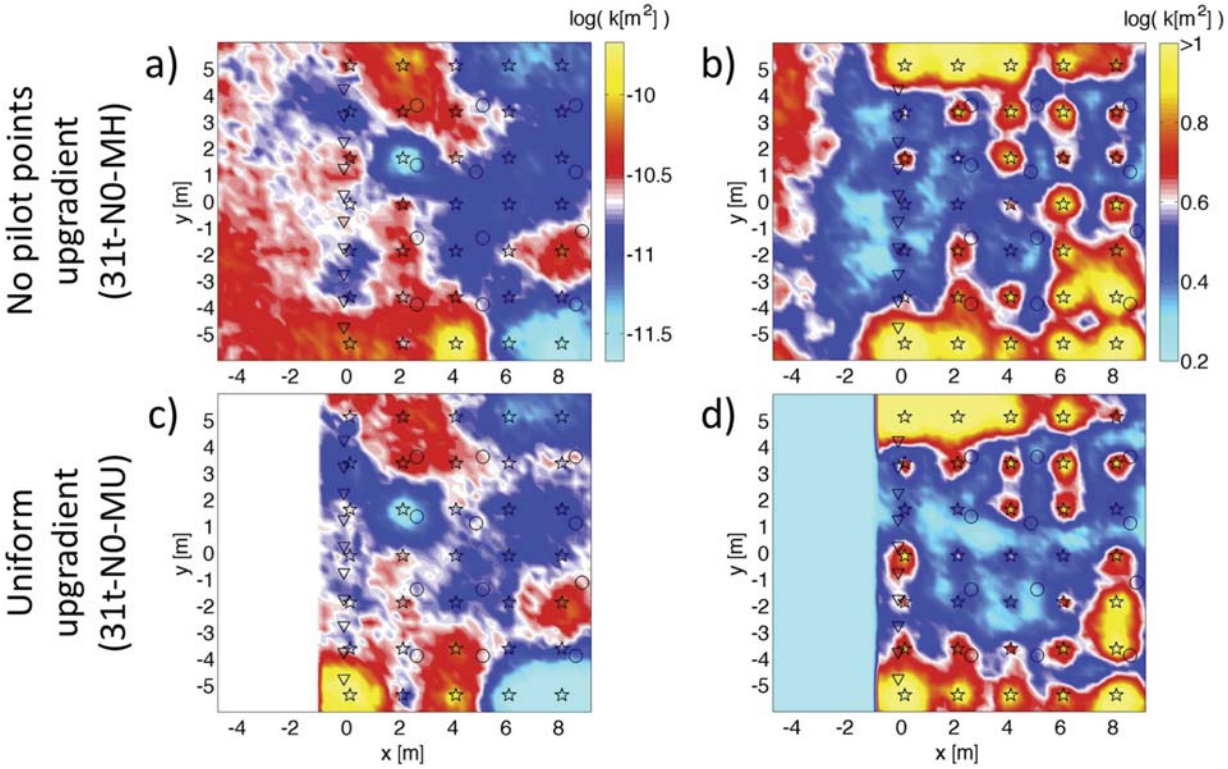

Figure 7. Sensitivity to the parameterization of region up gradient of (to the left of) the monitoring wells (see section 4.2 and Table 1). (a) The estimated log permeability and (b) the corresponding standard deviation for the case (31t-N0-MH), in which the up-gradient region is modeled with unconditional heterogeneity (i.e., no pilot points are used in that region). (c and d) Similarly, the same for the case (31t-N0-MU) in which the up-gradient region is assumed to be uniform and its value is estimated in the inversion.

(see Figures 5b–5d). For the former case, the log permeability error stays about the same (ε_1 decreases by 4%, while e_2 increases by 1%). For the case with high measurement error, ε_1 and ε_2 increase by 6% and 12%, respectively.

[49] Overall, the inversions continue to perform well despite the decreased number of sampling times and the addition of measurement error, though the case with less measurement error results in slightly more accurate permeability estimates.

4.4. Parameterization by Zonation With or Without Pilot Points (With 12 Sampling Times and No Measurement Error)

[50] As mentioned, to overcome nonuniqueness inherent to the inverse problem in this application, especially when extending the type of model from 2-D to 3-D, it may be beneficial to go beyond the geostatistical parameterization considered so far and reduce the number of unknown parameters, while taking advantage of geometrical information related to geological units that is potentially available through additional site characterization data, such as geophysical measurements, well logging data, or core descriptions. On the basis of the synthetic example discussed in sections 4.1–4.3, we examine the use of a zonation parameterization, where heterogeneity is described by zones within which hydrological properties are uniform and for which the spatial distribution is known, and examine decisions made in implementing the zonal parameterization in a model under different assumptions.

[51] We construct two synthetic "facies" data sets that represent the log permeability distributions with three and five types of zones, respectively, representing geological units or facies, each corresponding to a mutually exclusive range of log permeability. To construct each data set, for each pixel of the true log permeability distribution of the synthetic example (Figure 9a), a facies type is assigned

Figure 6. Sensitivity to pilot point placement and geostatistical errors (see section 4.1 and Table 1). The estimated log permeability for (a) coarse (case 31t-N0-C), (c) medium (case 31t-N0-M), (e) medium-shifted (case 31t-N0-M-Shifted), and (g) fine (case 31t-N0-F) spacing of pilot points. (b, d, f, h) The corresponding standard deviation. (i) The estimated log permeability and (j) standard deviation for case 31t-N0-M-Error in which medium spacing is used while the range is assumed to be twice as large as its true value of 3.3 m. Here and in the other figures showing permeability distributions (Figures 7, 8, 9, 12, and 16), the injection wells, monitoring wells, and pilot points are indicated with triangles, circles, and stars, respectively.

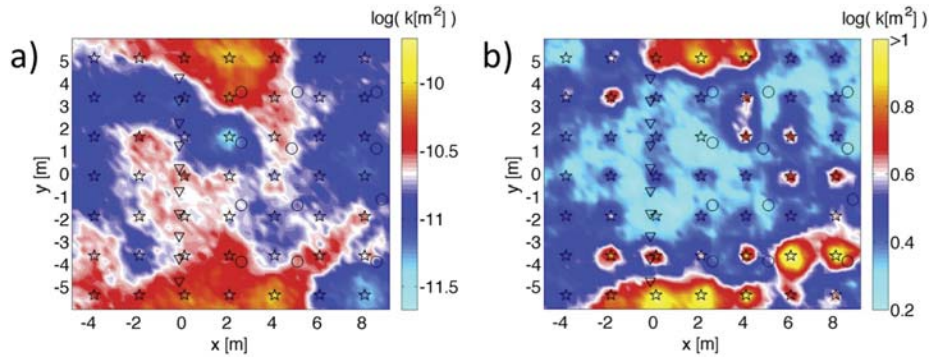

Figure 8. Inversion results for a representative case from section 4.3 (also see Table 2) in which the synthetic example is used to examine the sensitivity to decreased sampling times and increasing measurement error. (a) The estimated log permeability for the case with low noise (case 12t-N1-M) and (b) the corresponding standard deviation.

depending on its value of log permeability. In the first two cases, the facies data sets are error free and cover the entire model, consisting of either three zones (case 12t-N0-3Z; Figure 9b) or five zones (case 12t-N0-5Z; Figure 9c). In the third case (12t-N0-5Z-Error), the facies data set covers the entire model but contains errors, which were created intentionally by shifting 2 m by 2 m blocks by only 50 cm (or two grid blocks) in an arbitrary direction. In the last two cases, the facies data sets are known accurately, but they cover only a limited region that contains the wells, while the region outside of the wells is modeled either with an additional homogeneous zone (case 12t-N0-Zlim; Figure 9e) or with a geostatistical parameterization using pilot points (case 12t-N0-ZlimPP; Figure 9f). In each case, permeability values for each zone and a single porosity value are estimated by inversion. In the last case, 24 pilot point values and a reference log permeability value are also estimated. The synthetic concentration data used in these cases correspond to the coarse sampling scenario (t12) with no measurement error (N0). The inversion results are summarized in Table 3, and the estimated log permeability values are depicted in Figure 9.

[52] The purpose of the first two cases (12t-N0-3Z and 12t-N0-5Z) is to test whether such a parameterization can successfully represent heterogeneity under the most ideal conditions (i.e., in which facies data are perfectly known and are available for the entire model domain). Indeed, the inversion for the case with five zones (12t-N0-5Z) results in values of error variance ($s_0^2 = 0.051$) and log permeability error ($\varepsilon_1 = 0.172$, and $\varepsilon_2 = 0.044$) that are substantially lower than for the previous cases. However, for the case with only three zones (12t-N0-3Z), the results are poor (e.g., $s_0^2 = 0.767$) and indicate that using three zones does not meet the minimum requirements for representing heterogeneity at the site.

[53] The third case (12t-N0-5Z-Err), containing a very large value of s_0^2 (0.419) reveals that even relatively small errors in the zonal geometry can prevent the model response from being able to reproduce the (synthetic) measurements, and the resulting permeability parameter estimates are a poor representation of the true system.

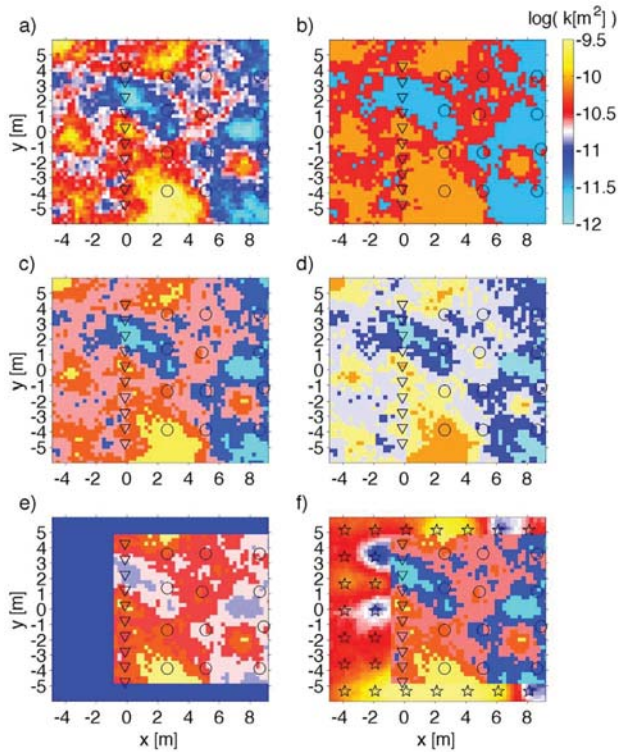

Figure 9. Inversion results for section 4.4 (see Table 3). (a) The true permeability. (b) The geometry of three zones is known perfectly (case 12t-N0-3Z). The geometry of five zones is known (c) perfectly (case 12t-N0-5Z), (d) inaccurately, with 2 m by 2 m regions shifted by ± 50 cm in the y direction (case 12t-N0-5Z-Error), and (e) perfectly, but only for the well region, where core data are available (case 12t-N0-5Zlim). (f) The zone geometry of five zones is only known in the well region, and pilot points are used in the region outside of the wells (case 12t-N0-ZlimPP). In all cases, permeability values in each zone along with a uniform porosity are estimated, while in the last case, 24 pilot point values are also estimated.

From this exercise, it is clear that uncertainty in the zonal geometry must be taken into account (see section 5.2).

[54] For the second to last case (12t-N0-Zlim), in which coverage of the facies data is limited to the well region, the error variance s_0^2 is dramatically increased (1.17), as is the error in the estimated log permeability ($\varepsilon_1 = 0.334$, and $\varepsilon_2 = 0.172$). This result is similar to the result of section 4.2, which showed that assuming a uniform zone up gradient of the injection wells could negatively affect the fit to the observations and the corresponding parameter estimates. But it serves as an important reminder that when zonal geometry is known for a subset of the model domain, the remaining regions of the model must be handled carefully.

[55] In the final case (12t-N0-ZlimPP) of the synthetic example, the zonation parameterization in the well region, based on the "limited" facies data, is combined with the geostatistical parameterization in the region outside of the wells, and the inversion results are greatly improved. The error variance s_0^2 is decreased to 0.224, and the error in log permeability is at a similar level as for the first case in this section with facies data available everywhere (case 883 12t-N0-5Z). However, this improvement comes at the expense of introducing an additional 24 parameters to be estimated by inversion.

[56] We conclude that even relatively small uncertainty in the facies geometry must be taken into account to avoid biased parameter estimates. In addition, even if perfect information on the spatial distribution of facies within the well region is known, making the decision to assume homogeneous properties outside of the well region is insufficient for obtaining accurate parameter estimates. However, combining such information with a pilot point parameterization in the remaining region appears to work well. A similar approach to the one considered in this section will be applied to the field data in section 5.2 to estimate 3-D properties.

5. Application to Field Data

[57] Next we perform hydrological inversion of the field data collected in the Winchester experiment, while taking advantage of the lessons learned in the synthetic example of section 4. We use a 2-D model with the geostatistical parameterization in section 5.1. Then, we extend the model to 3-D by introducing facies data into the inversion procedure through a combined application of the zonation parameterization and the geostatistical parameterization in section 5.2. Details of the corresponding inversion cases and results are given in Tables 4 and 5.

[58] The bromide concentration data that were measured in the experiment (Figure 10) are used for the remaining cases. Recall that the synthetic data of the coarse sampling scenario 12t of section 4 (Figures 5b–5d) were intentionally made similar to the actual field data.

5.1. Results for 2-D Model With Pilot Point Parameterization

[59] In cases Field-M and Field-F, we apply the geostatistical parameterization with medium (case Field-M) and fine (case Field-F) pilot point spacing, which has 49 and 80 pilot points, respectively, as unknown parameters, along with a reference value of log permeability and the porosity. These pilot point configurations correspond to the medium and fine spacing configurations used in section 4, for example, in cases 31t-N0-M and 31t-N0-F.

[60] The fit between the measured and simulated concentrations are shown in Figure 10, and the residuals (the

Table 4. Summary of 2-D Inversions for Field Data Application

	Case			
	Field-M	Field-F	Field-M-10deg	Field-M-VarPor
Section	5.1	5.1	5.1	5.1
Description	Field data, medium pilot point spacing	Field data, fine pilot point spacing	Field data, medium pilot point spacing, gradient rotated by 10°	Field data, medium pilot point spacing, variable porosity
Number of pilot points m_{pp}	0	0	0	0
Number of observation times	12	12	12	12
Number of bromide data m_b	144	144	144	144
Measurement error σ_{bromide}[a] (mmol)	0.13	0.13	0.13	0.13
Number of pilot points m_{pp}[b]	49	80	49	49
Number of unknowns n	51	82	51	52
$m_b + m_{pp} - n$ (equation (A1))	142	142	142	141
Parameter estimates				
Pilot points	Figures 12a and 12b	Figures 12c and 12d	Figures 12e and 12f	Figures 12g and 12h
$\sigma_{\log k,ave}$ (equation (B3))	0.434	0.436	0.403	0.437
$\log k_{ref}$	−10.626 (±0.033)[c]	−10.620 (±0.013)	−10.635 (±0.024)	−10.638 (±0.031)
Uniform porosity ϕ	0.121 (±0.012)	0.120 (±0.011)	0.117 (±0.010)	N/A
Variable porosity $\phi = a\,(K\,(\text{m d}^{-1}))^{1/b}$	N/A	N/A	N/A	$a = 0.099$ (±0.007), $b = 13.1$ (±5.6)
Residual analysis				
Error variance s_0^2 (equation (A1))	0.525 (±0.084)	0.573 (±0.079)	0.538 (±0.064)	0.528 (±0.087)
NS (equation (A2))[d]	0.916 (±0.015)	0.922 (±0.008)	0.923 (±0.010)	0.915 (±0.014)

[a]Standard deviation measurement error assumed for the field data in section 5.
[b]The pilot point parameters (log permeability values) are given prior data equal to log k_{ref}, with a standard deviation of 1.0.
[c]Average value for all realizations followed in parentheses by the standard deviation.
[d]Nash-Sutcliffe model efficiency.

Table 5. Summary of 3-D Inversion for Field Data Application

	Field-ZlimPP Case
Section	5.2
Description	Five zones from facies model; pilot points outside of wells
Region with lithofacies zones	-1 m $< x <$ 12 m, -5 m $< y <$ 5 m
Region with pilot points	For $x < -1$ m, all y; for $x > -1$ m, $y < -5$ m, $y > 5$ m
Number of observation times	12
Number of bromide data m_b	144
Measurement error σ_{bromide}[a] (mmol)	0.13
Number of pilot points m_{pp}[b]	24
Number of unknowns n	29
$m_b + m_{pp} - n$ (equation (A1))	139
Parameter estimates	
Pilot points (PP)	Figure 16
$\log k_{ref}$	-10.73 (± 0.09)[c]
$\log k_1$	fixed
$\log k_2$	-9.89 (± 0.35)
$\log k_3$	-10.69 (± 0.09)
$\log k_4$	-10.96 (± 0.13)
$\log k_5$	fixed
ϕ	0.104 (± 0.006)
Residual analysis	
Error variance s_0^2 (equation (A1))	0.742 (± 0.089)
NS (equation (A2))[d]	0.869 (± 0.016)

[a]Standard deviation measurement error assumed for the field data in section 5.

[b]The pilot point parameters (log permeability values) are given prior data equal to $\log k_{ref}$, with a standard deviation of 1.0.

[c]Average value for all realizations followed in parentheses by the standard deviation.

[d]Nash-Sutcliffe model efficiency.

measured minus the simulated concentrations) are shown in Figure 11. The residual distributions appear to be mostly random and Gaussian, exhibiting little bias. The subtle effects of potential outliers can be seen in the residuals for several monitoring wells. For example, in M1 and M7 the tail (Figure 11) is slightly increased to the right because of a couple of measurements that are underpredicted by the simulated values (see Figure 10). Similarly, the tail is increased slightly to the left at well M9 because of two lower concentration measurements that are overpredicted in the simulations.

[61] The estimated log permeability and standard deviation are shown in Figure 12. The log permeability estimates are similar for both cases (Figures 12a and 12c), as is the porosity, and the uncertainty, expressed by the average standard deviation $\sigma_{\log k,ave}$ (Table 4).

[62] Next we compare the permeability estimates with values derived from slug test data collected at the site before and after the biostimulation experiment. For this purpose, the permeability with grid block centers within 25 cm of the wells was averaged using the geometric mean. In addition, corrections were made to account for the actual thickness of the aquifer at each well (the effective log permeability values estimated through the hydrological inversion correspond to an aquifer with 2.5 m aquifer thickness). As seen in Figure 13, the slug test values are consistently lower than the permeability estimates, by around a half order of magnitude in the monitoring wells, and by up to 2 orders of magnitude in the injection wells. The relatively poor match between the slug test estimates and the inversion estimates could be to incorrect assumptions in the slug test analysis (e.g., errors in the anisotropy value, which was assumed to be 0.1, or the specified radii of the near-well materials) or in the conceptual model used for the hydrological inversion (e.g., that the porosity is assumed constant, or errors in the assumed direction of regional groundwater flow).

[63] To investigate whether errors in the assumed direction of groundwater flow could bias the inversion results, we performed inversions with the groundwater flow direction either fixed or considered as an unknown parameter, thus allowing it to vary from the direction perpendicular to the line of injection wells. This is accomplished by fixing the pressure on all boundaries of the model, as opposed to only on the left and right sides of the model (Figure 1), and forcing it to satisfy the equation of a plane that is rotated by α. The angle α is relative to the line that is perpendicular to the injection wells and increases in the counter clockwise direction, such that $\alpha = 0°$ and 90° correspond to the gradient pointing from left to right and from bottom to top, respectively (Figure 1). While, technically, the angle α can either be fixed or considered as an additional parameter to be estimated by inversion, our investigation indicated that α cannot be uniquely determined along with the remaining parameters, as its estimated value appears to be quite sensitive to the initial guess of its value.

[64] In case Field-M-10deg, the regional gradient is assumed to be 10° different from the previous cases (i.e., α is fixed to 10° for the inversion instead of 0° for the previous cases). The fit to the concentration data remains similar to the previous cases Field-M and Field-F, while the average standard deviation of the estimated log permeability ($\sigma_{\log k,ave}$) is lower (decreased from 0.434 and 0.436 to 0.403), indicating slightly lower uncertainty. As can be seen in Figure 12e, the estimated permeability values are qualitatively very similar to the previous cases as well. The most obvious impact of rotating the gradient seems to be a decrease in the estimated log permeability around half of the injection wells ($-6 < y < -1$). However, the permeability estimates for this case do not show a significantly different match when compared to the slug test data (Figure 13).

[65] Until now, we assumed that the porosity was constant and estimated its value along with the other hydrological parameters, but it is not clear what impact this decision ultimately has on the estimated permeability values. To examine this issue, we included an additional case in which the porosity was made variable (heterogeneous), assumed to vary as a function of the permeability. For this purpose we chose a general form of the Bretjinksi model [see *de Marsily*, 1986] that was developed for sands: $\phi = a\ \mathrm{K}^{1/b}$, where ϕ is the porosity, K is the hydraulic conductivity (m d^{-1}), and the empirically determined factors $a = 0.117$ and $b = 7$ (see Figure 14). We implemented this relationship in TOUGH2 to make the porosity a function of permeability for case Field-M-VarPor. In addition to the unknown pilot points, we estimated the parameters a and b as unknown parameters in the inversion, along with the remaining parameters (see Table 4). The resulting estimated permeability distribution and standard deviation are shown in Figures 12g and 12h, respectively, and the estimated porosity-permeability functions are shown in Figure 14. The remarkable similarity between the cases Field-M and Field-M-VarPor of the parameter estimates, the error variance, and the other performance measures indicates

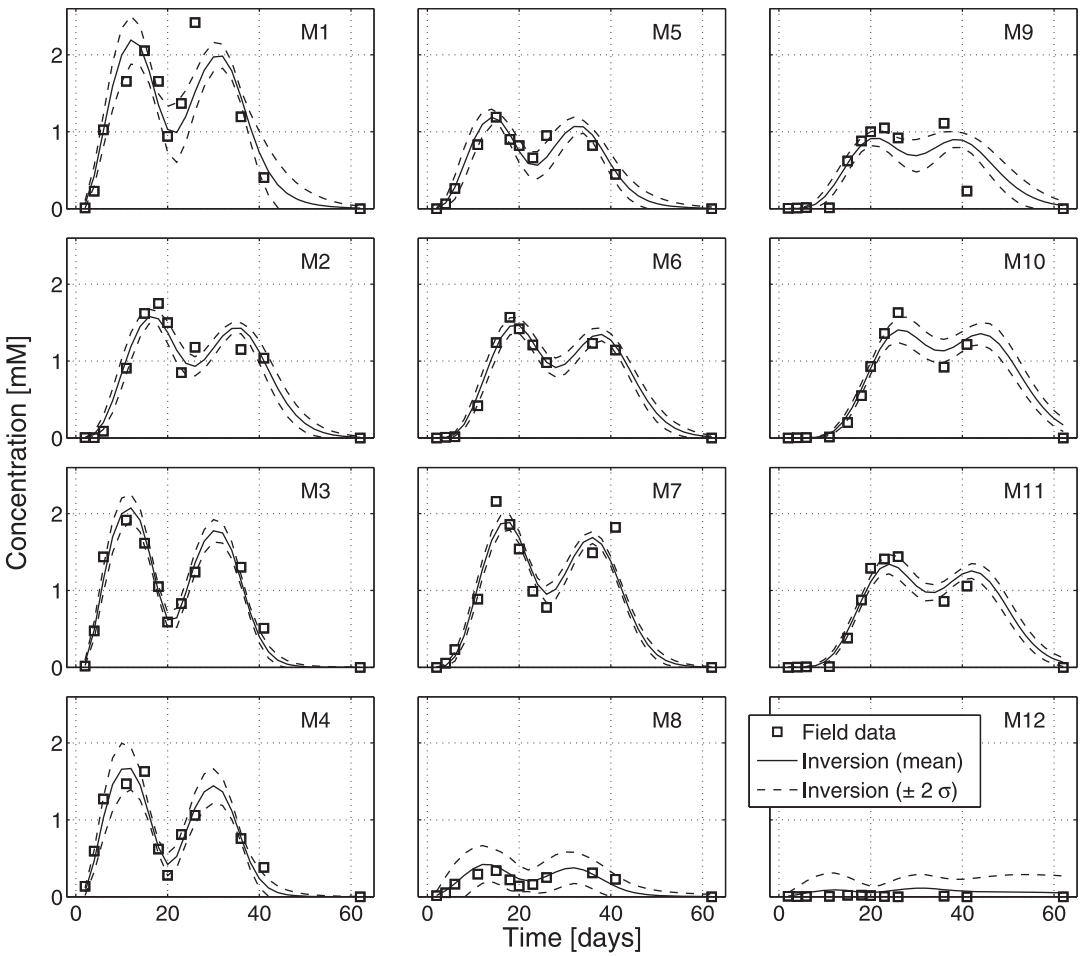

Figure 10. Measured and simulated concentrations for the inversion of field data in case Field-M of section 5.1 (see Table 4). For monitoring wells M1 to M12, the field data are shown with symbols, and the inversion results are shown with solid lines (mean of 30 realizations) and dashed lines (mean ±2 standard deviations).

that the variations in porosity, at least as it was implemented, does not significantly impact the parameter estimates and may not be significant for transport.

[66] Comparison of the estimated porosity values in sections 4 and 5 to measured values from the site is made difficult due to a lack of direct porosity measurements at the site (e.g., the recovered core at the site is not representative of the in situ conditions as the largest fraction could not be collected after coring). So porosity estimates must be inferred indirectly. For example, despite the uncertainty in the contents of the cored material, *Yabusaki et al.* [2011] used a packing model from the measured grain size distributions from 23 samples to estimate porosity for the representative facies (described further in section 5.2), obtaining values between 0.2 and 0.23, with an overall mean of 0.224. Electrical conductivity logging was also performed with a Geonics EM-39 borehole sonde tool in most of the wells in the experimental plot, but translation into porosity requires knowledge of the fluid electrical conductivity, and parameters of a function relating bulk electrical conductivity to porosity [*Archie*, 1942]. Assuming a value for the fluid electrical conductivity (2700 μS cm^{-1}) based on a recent measurement at the site and assuming a range of possible values of the cementation factor m [see *Archie*, 1942] provide a possible range for the porosity at the site (Figure 15). For a value of m between 1.3 and 1.8, which might be expected for unconsolidated sediments, the average porosity may range between 0.1 and 0.2, with a standard deviation between 0.04 and 0.05, which is entirely consistent with the result shown in Figure 14, and the values of porosity estimated in previous cases in this study (around 0.12).

5.2. Results for 3-D Model With Facies-Based Zonation and Pilot Point Parameterization

[67] So far in this study, it was assumed that a 2-D representation of heterogeneity in the hydrological model was sufficient, and the effects of vertical heterogeneity were lumped into an effective (depth-averaged) permeability that varied in the horizontal direction. The choice of a 2-D model over a 3-D model can be justified for at least two reasons. First, the thickness of the actual aquifer is small compared to the horizontal dimensions of the model domain, which minimizes the importance of accounting

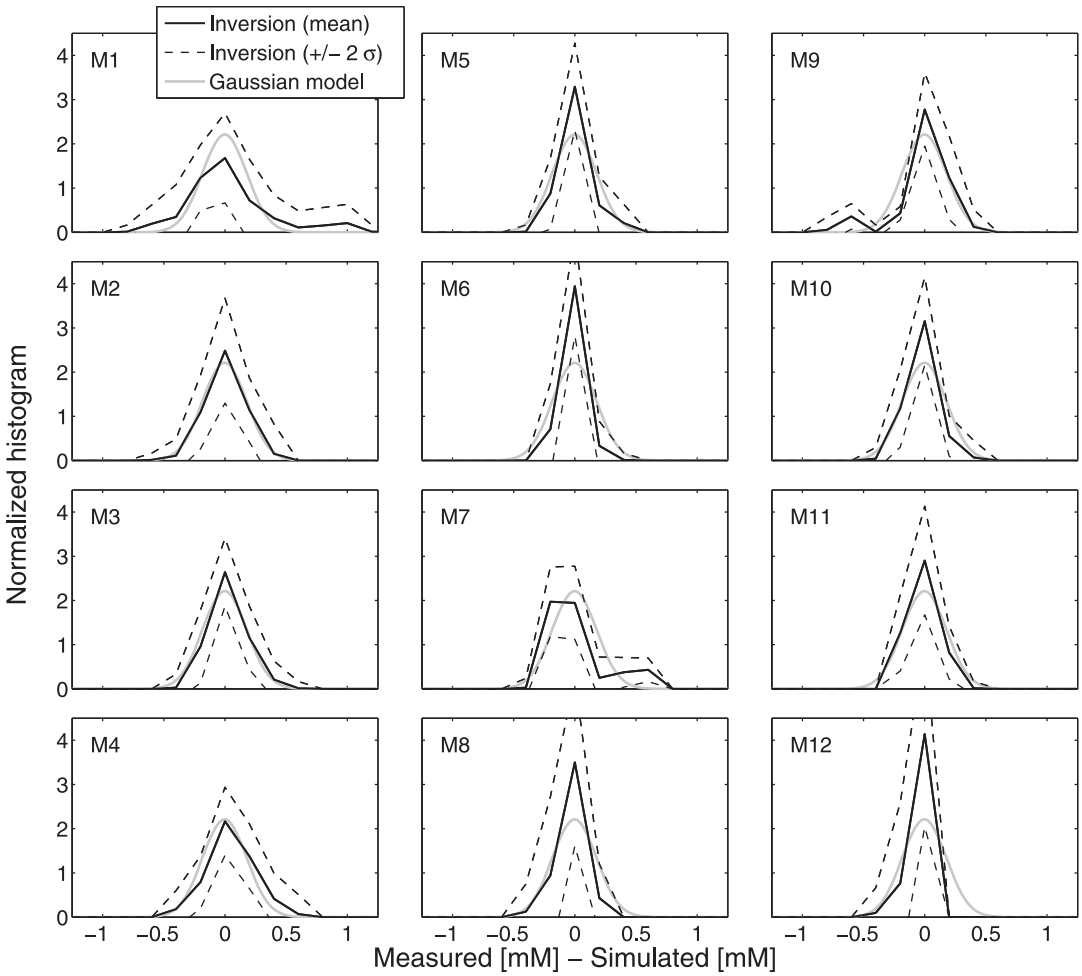

Figure 11. Residuals (difference between measured and simulated concentrations) for the inversion of field data in case Field-M of section 5.1 (see Table 4). For monitoring wells M1 to M12, the normalized histograms are shown with solid lines (mean of 30 realizations) and dashed lines (mean ±2 standard deviations), and the corresponding Gaussian model is shown with a gray line.

for flow variations in the vertical direction. Second, since the available data do not contain information about vertical flow variations (as they are depth-averaged concentrations), there is insufficient information to parameterize a 3-D geostatistical model with pilot points (i.e., developing such a model would lead to nonuniqueness and substantially increased estimation uncertainty and is thus not justifiable).

[68] While the complexity of the 2-D model and its parameterization does appear sufficient to reproduce the measured data (as was demonstrated in section 5.1), it is known from geologist's log descriptions that the site exhibits vertical heterogeneity, and there is interest in better understanding the effect of vertical variations on flow and transport, and on biostimulation in particular. Extending the 2-D model to 3-D may be possible using facies information available from geologic data [*Yabusaki et al.*, 2011].

[69] In this section, we benefit from multiple realizations of facies models previously developed on the basis of a geostatistical analysis of geologic data (see *Yabusaki et al.* [2011] for details). For a given model, the 3-D facies distribution is known, and each pixel is assigned a facies indicator. In total, five facies indicators are considered: 1 for the clay fill layer, 2 for fines, 3 for muddy gravel, 4 for sandy gravel, and 5 for the Wasatch bedrock formation. In the modeling study of *Yabusaki et al.* [2011], these models were used to parameterize a 3-D reactive transport model for the Big Rusty experiment of 2008. For this purpose, the following values of hydraulic conductivity were assumed for each facies (the base 10 log of permeability (m^2) is given in parentheses for reference): for facies 1, 0.1 m d^{-1} (−12.9); for facies 2, 0.01 m d^{-1} (−13.9); for facies 3, 3.0 m d^{-1} (−11.45); for facies 4, 30.0 m d^{-1} (−10.45); for facies 5, 1×10^{-5} m d^{-1} (−16.9).

[70] In the final inversion case for this study, we combine the 3-D zonation parameterization for the region covered by facies data (-1 m $< x <$ 12 m and -5 m $< y <$ 5 m and for all values of z) and a 2-D geostatistical parameterization for the remaining region (for $x < -1$ m, all values of y; for $x > -1$, $y < -5$ m or $y > 5$). In other words, the area covered by the facies data is vertically heterogeneous and is described by five zones with a single permeability value

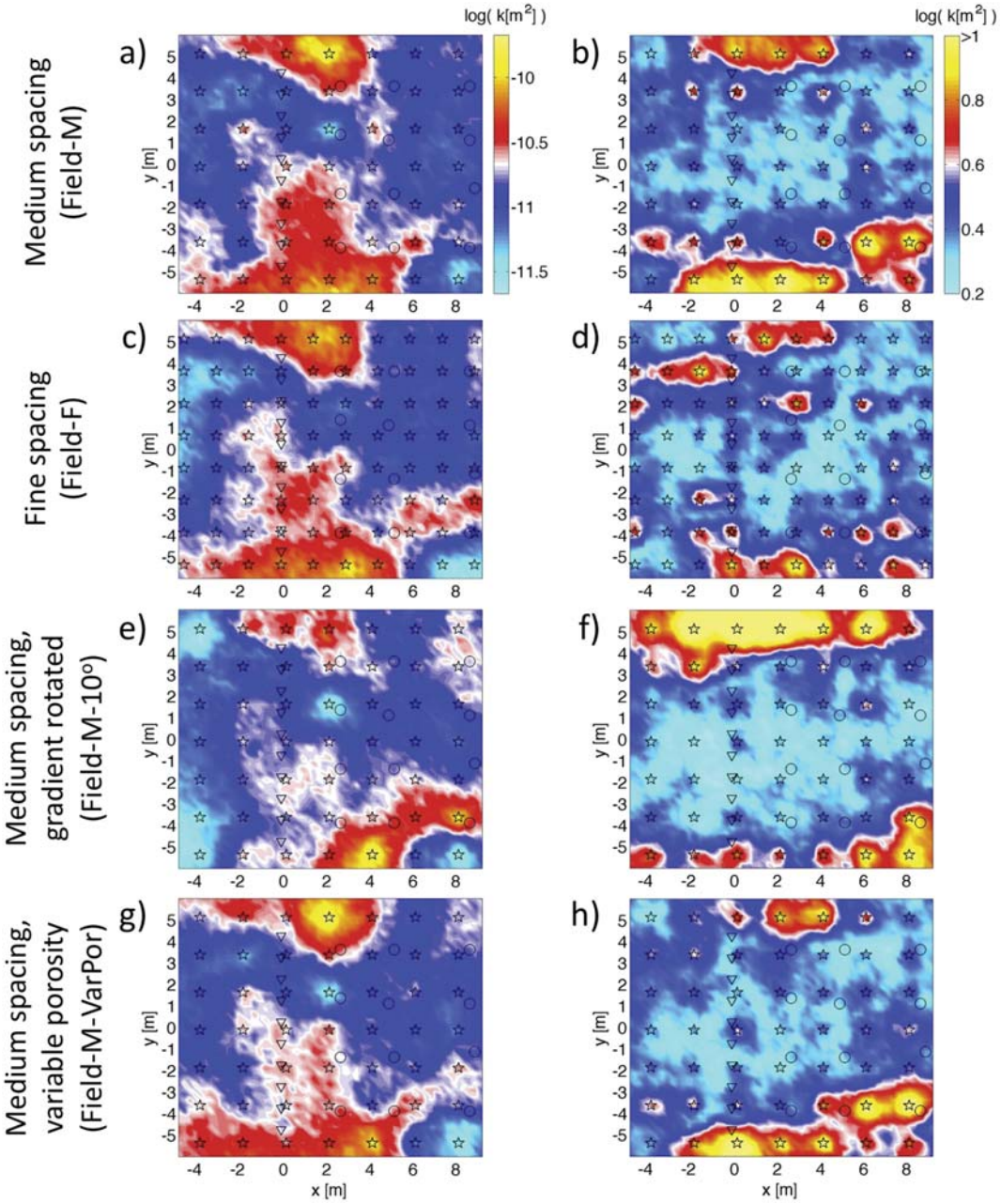

Figure 12. Results for the 2-D inversion of field data (see section 5.1 and Table 4). The estimated log permeability for the cases with (a) medium (case Field-M) and (c) fine (case Field-F) pilot point spacing are shown. (e) Case Field-M-10°, which assumes the gradient is rotated by 10° counterclockwise. (g) Case Field-M-VarPor, which assumes variable porosity and medium pilot point spacing. (b, d, f, h) The standard deviation for these cases, respectively.

each; for the area surrounding the facies data, the permeability is uniform in the vertical direction but heterogeneous in the horizontal direction, described by the geostatistical parameterization using 24 pilot points (with the same pilot point configuration as was given in section 4.4). In total, the log permeabilities of 3 of the facies zones are estimated during inversion, in addition to the 24 pilot point values, the corresponding log k_{ref}, and the porosity. (The permeability in the bedrock is fixed at the value listed above, as there is insufficient sensitivity in the data to estimate its value. In addition, the clay fill layer is not present at the depths covered by the 3-D model, so it is not included.) The inversion process is repeated for 30 realizations, each using a different facies model realization, and a different seed number for generating the log permeability in the geostatistical region containing pilot points.

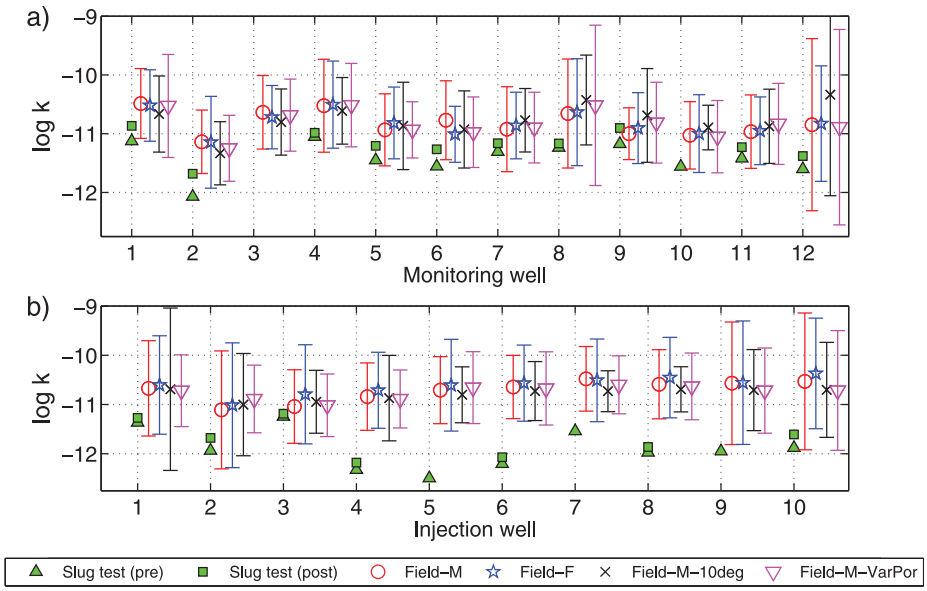

Figure 13. Permeability values estimated in 2-D inversion cases of section 5.1, as indicated in the legend, compared with values inferred from slug test data before (pre) and after (post) the experiment for the (a) monitoring wells and (b) injection wells. The monitoring and injection well locations are shown in Figure 1 (the injection wells are numbered from top to bottom). Permeability values within 25 cm of the wells are averaged with geometric averaging.

[71] Details of the inversions are given in Table 5. The log permeability for a single inversion realization and the mean of all 30 realizations are shown in Figures 16a and 16b, respectively. While the error variance is somewhat increased ($s_0^2 = 0.74$) relative to the cases in section 5.1 (0.53–0.57), it is nonetheless a remarkable fit, evidenced by the still large value of NS (0.869), especially considering the reduced number of parameters required for fitting the data.

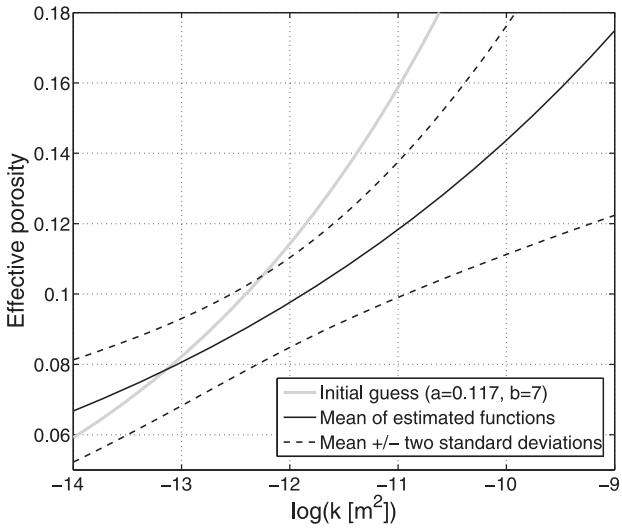

Figure 14. Mean of the estimated porosity-permeability relationships obtained by inversion with medium pilot point spacing and an assumed relationship between porosity and permeability (case Field-M-VarPor).

[72] The estimated values of log permeability for facies 2, 3, and 4 are −9.9, −10.7, and −11.0, respectively. While the values for facies 3 and 4 are similar to the previously assumed values, the value for facies 2 is significantly higher.

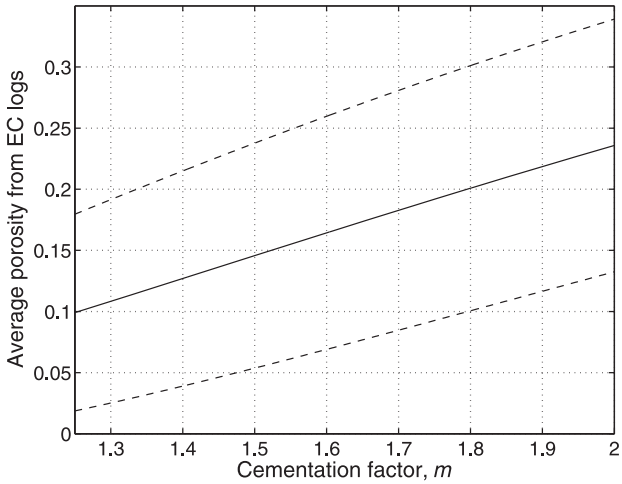

Figure 15. Average porosity inferred from electrical conductivity logs as a function of the cementation factor m in Archie's law [*Archie*, 1942], which is used to convert the bulk electrical conductivity to porosity. The solid line shows the average value, while the dashed lines represent the average value ±2 standard deviations. The fluid electrical conductivity was set to 2700 μS cm^{-1} on the basis of a recent measurement of fluid electrical conductivity at the site collected during the same time of season as the Winchester experiment.

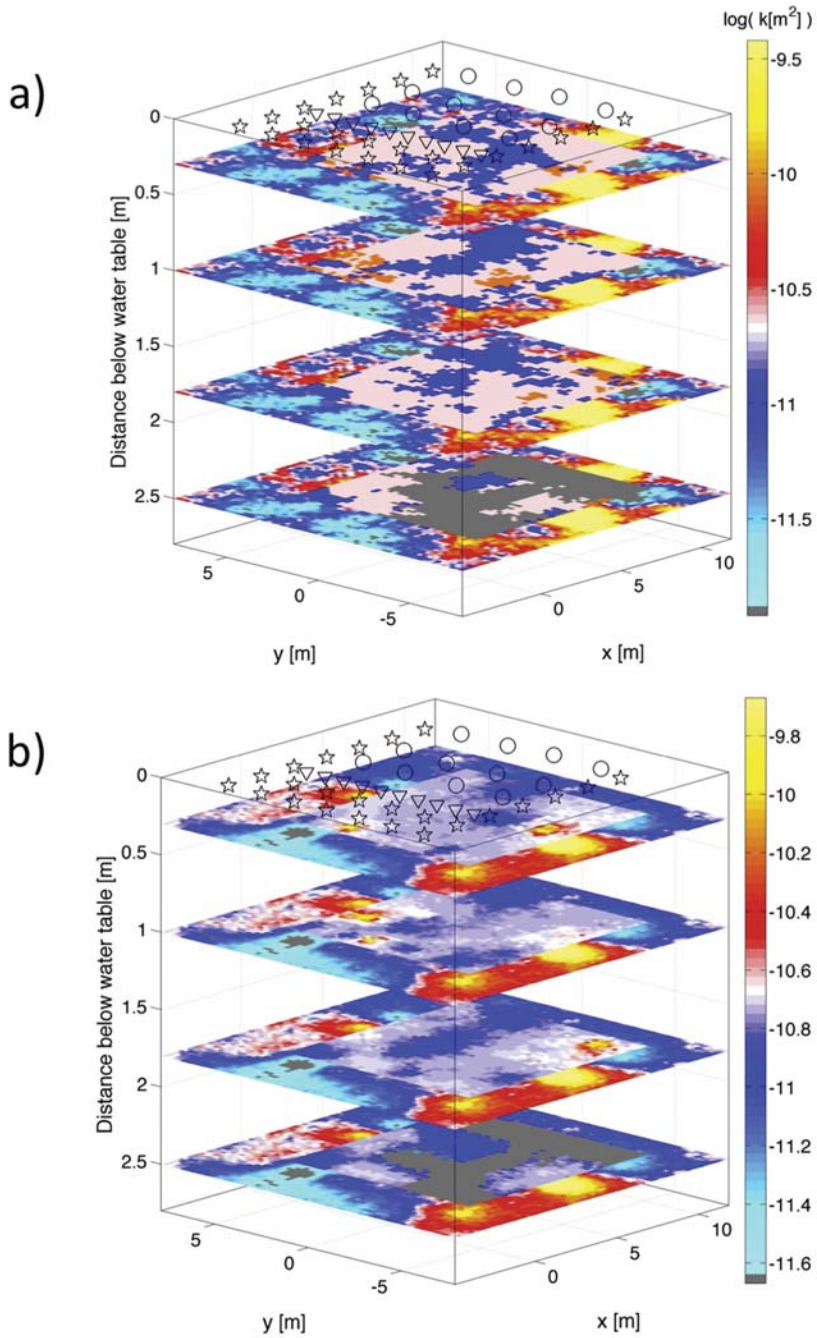

Figure 16. Distributions of log permeability for 3-D models of section 5.2 (a) obtained by a single inversion realization and (b) obtained by taking the mean of 30 inversion realizations. Note that the color scales are different, and gray colors indicate values lower than the ranges of the color scales.

[73] No physical explanation is evident for how the estimate of log permeability for the fines could be higher than for the muddy gravel and sandy gravel units, but it may result from the limited occurrence of facies 2 in the model. For all 30 realizations, the average percentages of volume corresponding to facies 2, 3, 4, and 5 in the model region covered by facies data are 6.1, 53.2, 25.9, and 14.8, respectively. (Recall that the permeability value of facies 5 is fixed in the inversion.) For the entire model, including the pilot point region, the corresponding average percentages are 2.4, 21.0, 10.2, and 5.8. The limited occurrence of facies 2 leads to decreased sensitivity of the measurements to its permeability value, which may explain its higher estimation uncertainty (σ is 0.35 for facies 2, as opposed to 0.09 and 0.13 for facies 3 and 4, respectively) and higher than expected estimated value. While not attempted here, one could constrain the

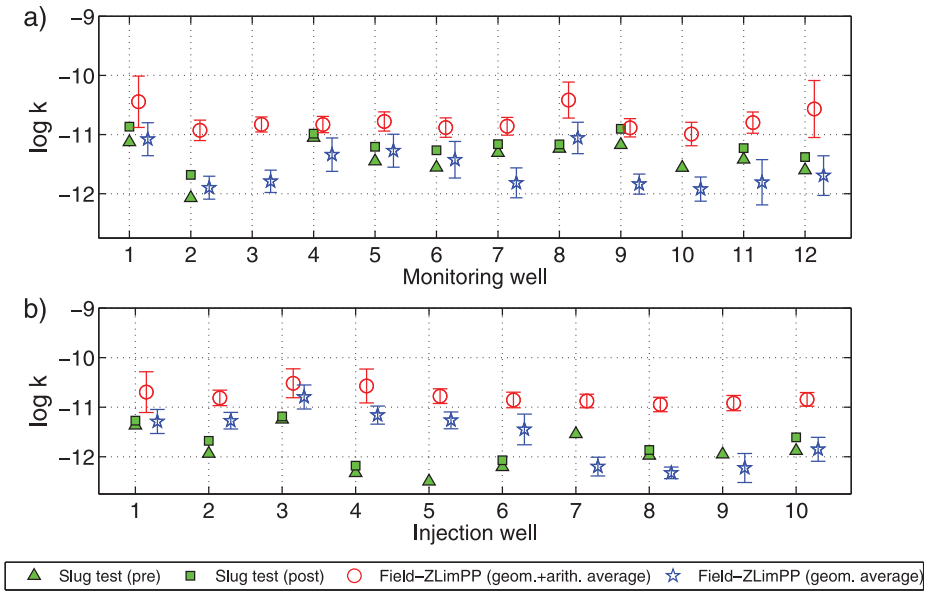

Figure 17. Vertical average of the permeability values that were estimated in the 3-D inversion of section 5.2 (case Field-ZLimPP) compared with values inferred from slug test data before (pre) and after (post) the experiment for the (a) monitoring wells and (b) injection wells. The monitoring and injection well locations are shown in Figure 1 (the injection wells are numbered from top to bottom). Permeability values within 50 cm of the wells are averaged in one of two ways: with geometric averaging in the horizontal direction followed by arithmetic averaging in the vertical direction (circles) or with geometric averaging (stars).

inverse problem to enforce the rank of permeability values among facies, assuming such rank were known with confidence.

[74] Permeability values from the 3-D models within 25 cm of the well locations were averaged for comparison with the slug test data (Figure 17). Two forms of averaging were done: one in which the nearby values at each depth were horizontally averaged with the geometric mean, followed by averaging of the resulting values in the vertical direction with the arithmetic mean; and one in which the geometric mean of all the nearby values was used (regardless of direction). The case with the geometric and arithmetic averaging results in values that are quite similar to the values obtained in the 2-D case (compare Figures 13 and 17). However, the 3-D values that were averaged with the geometric mean are lower, and the match to the slug test data is closer in some areas. These results suggest that the slug test measurements could represent a different averaging of permeability than is represented by the depth-averaged estimates obtained by hydrological inversions with the 2-D model. Properly determining how near-well heterogeneity is averaged using different aquifer tests is possible with additional effort [*Wu et al.*, 2005]. Current research involving high-resolution modeling of the slug tests at the Rifle site is underway and will hopefully provide some guidance on how to interpret such data in the future at this site and similar sites.

6. Conclusions

[75] We consider the estimation of spatial variations in permeability and several other parameters through inverse modeling of tracer data, specifically synthetic and actual field data associated with the 2007 Winchester experiment from the DOE Rifle site. This site represents a challenging setting that reflects the real-world complexities associated with dynamic, shallow, alluvial aquifers. Our aim is to highlight and quantify the impact on inversion results of various decisions related to parameterization, such as the positioning of pilot points in a geostatistical parameterization, the handling of up-gradient regions, the inclusion of zonal information derived from geophysical data or core logs, extension from 2-D to 3-D, assumptions regarding the gradient direction, porosity, and the semivariogram function, and deteriorating experimental conditions.

[76] A synthetic example that was based on the field experiment allowed for the impact of subtle changes in pilot point alignment and spacing to be evaluated. The homogenous inversions were performed to provide an idea of the extent to which the measured concentrations were affected by heterogeneity versus an irregular injection function. Moving from the coarse to medium spacing pilot point configuration (essentially decreasing the spacing by only one grid block) caused a fair improvement but a negligible improvement was found with further reduced spacing (decreasing the spacing by one more grid block). Shifting the pilot point networks uniformly by several grid blocks did not change the results significantly. However, the configuration in which the pilot points were not aligned horizontally did not perform as well as the regular spacing, suggesting that uniform spacing is better in this type of application and environment. Intentionally increasing the integral scale by a factor of 2 did not result in significant deterioration of the results, which means that some

uncertainty in the geostatistical model due to a lack of site-specific data can be allowed. We attempted to estimate the parameter in the inversion but were unable to do so successfully; interestingly, the optimization algorithm tended to increase the integral scale, resulting in a better fit between the measured and simulated concentrations but worse parameter estimates.

[77] The handling of the parameterization in the up-gradient region was far more important than the pilot point spacing. The measured concentrations could not be adequately simulated with a uniform zone or with unconditioned heterogeneity in the up-gradient region. Adding up-gradient pilot points has the obvious disadvantage of necessitating an increased number of parameters for inverse modeling.

[78] We also examined parameterization by zonation in order to reduce the number of unknown parameters by taking advantage of geometrical information related to geological units, even if it is only available for a subset of the model domain. In the most ideal case in which the facies distribution is known with five zones over the entire model, the permeability estimates are improved relative to the geostatistical parameterization. However, results are poor when only three zones are used. Further demonstrating the sensitivity of the results to details of the zonation information, when the five zones are assumed to be known everywhere but slightly inaccurate, the performance declines severely. In practical application, the success depends heavily on the resolution and accuracy of the facies model. However, with multiple realizations of facies, for example, generated from geostatistical simulation (see section 5.2 and *Yabusaki et al.* 2011]), the uncertainty in zone geometry is naturally taken into account. It was also demonstrated that the extent of the facies zone is crucial (i.e., the up-gradient boundary could not be modeled with uniform properties, but a combination of pilot points and the zonation information worked well).

[79] Inversion of the actual field data was performed first for a geostatistical parameterization of the permeability using a 2-D model. Residuals between the measured and simulated concentrations seem roughly normally distributed. The match does not improve when inversion is performed with the assumption of a different gradient direction or even when the porosity is allowed to be variable. The assumption that the porosity is uniform for this study seems reasonable so far. The slug test–derived estimates of log permeability are about 0.5 or 1 order of magnitude lower than those estimates in the inversion. Inversion was also extended to 3-D by combining the 3-D facies distributions with 2-D pilot points. In that case, the depth-averaged permeability estimates are similar to the 2-D results when a geometric average in the horizontal direction followed by an arithmetic average in the vertical direction is performed. However, the match to the slug test data is better in some regions when a geometric mean of the permeability in the 3-D model is used, indicating that the slug test measurements could represent a different averaging of permeability than is represented by the depth-averaged estimates obtained by hydrological inversions with the 2-D model.

[80] Properly determining how near-well heterogeneity is averaged using different aquifer tests is possible with additional effort [*Wu et al.*, 2005]. Current research involving high-resolution modeling of the slug tests at the Rifle site is underway and will hopefully provide some guidance on how to interpret such data in the future at this site and similar sites.

[81] This study adds to the relatively limited number of studies that offer guidance on the use of pilot points in complex real-world experiments involving tracer data (as opposed to hydraulic head data). It highlights the importance of proper spatial parameterization of subsurface heterogeneity, as errors in the model structure are partly compensated for by estimating biased property values during the inversion of tracer data. These biased estimates, while potentially providing an improved fit to the calibration data, may lead to wrong interpretations and conclusions (e.g., regarding the impacts of biostimulation on flow and transport properties) and reduce the ability of the model to make reliable predictions. Influences of heterogeneity outside the immediate zone of interest cannot be ignored, and proper attention has to be paid to the averaging scheme employed when reducing the model dimension. We demonstrated a combined approach that allows the modeler to flexibly include available information as deterministically as possible, while using a stochastic method to account for uncertainty where such information is not available.

Appendix A: Residual Analysis

[82] To quantify how well model output reproduces measured data for each inversion case, we analyze the residuals (the difference between the measured and modeled output) using two statistical measures: the error variance s_0^2 and the Nash-Sutcliffe model efficiency (NS).

[83] The a posteriori error variance is the variance of the weighted residuals given by

$$S_0^2 = \frac{1}{m_b + m_{pp} - n} \sum_{i=1}^{m_b+m_{pp}} \left(\frac{d_i - z_i}{\sigma_i} \right)^2, \tag{A1}$$

where d_i and z_i are the measured and simulated values of observation i, σ_i is the standard deviation of the measurement error, m_b is the number of bromide concentration data, m_{pp} is the number of prior data (associated with the pilot points), and n is the number of parameters estimated in the inversion. The quantity $(m_b + m_{pp} - n)$ is the degree of freedom of the inversion.

[84] A popular method for comparing model performance is given by the NS criterion, which is a measure of the ratio of model error to variability in the measured data:

$$\mathrm{NS} = 1 - \sum_{i=1}^{m_b} (d_i - z_i)^2 \Big/ \sum_{i=1}^{m_b} (d_i - \overline{d})^2, \tag{A2}$$

where $\overline{d}$ is the mean value of all m_b bromide concentration data. The NS index, which is a number less than or equal to one, can be interpreted as the relative ability of a model to predict the data, where NS = 0 indicates that the model is not better at predicting the data than simply obtaining the mean of the observed values. Performance levels for the NS index have been specified as follows [*Maréchal*, 2004; *Allen et al.*, 2007]: >0.65, excellent; between 0.5 and

0.65, very good; between 0.2 and 0.5, good; <0.2, poor. *Gupta et al.* [2009] provide a detailed interpretation of the NS criterion.

Appendix B: Measures of Parameter Uncertainty and Estimation Error

[85] Several measures are used in this study for quantifying uncertainty in the estimated log permeability (log k). The mean log k estimate for N inversion realizations is calculated at each grid block i as follows:

$$\langle \log k_{est,i} \rangle = \frac{1}{N} \sum_{j=1}^{N} \log k_{est,i}^{j}, \tag{B1}$$

where the superscript j is the realization number. The standard deviation gives a measure of the corresponding estimation uncertainty:

$$\sigma_{\log k,i} = \sqrt{\frac{1}{N} \sum_{j=1}^{N} \left(\log k_{est,i}^{j} - \langle \log k_{est,i} \rangle\right)^{2}}, \tag{B2}$$

the average standard deviation for all m grid blocks in a region of interest is given by

$$\sigma_{\log k,ave} = \frac{1}{M} \sum_{i=1}^{M} \sigma_{\log k,i}, \tag{B3}$$

where i is grid block number. This quantity is reported in Tables 1, 2, and 4.

[86] For cases in section 4 involving a synthetic example in which the true log k is known, we also report measures of the average estimation error. One such measure is the average of the absolute value of the difference between the true log k values and the mean of the estimated log k values as follows:

$$\varepsilon_1 = \frac{1}{M} \sum_{i=1}^{M} \left| \log k_{true,i} - \langle \log k_{est,i} \rangle \right|. \tag{B4}$$

A second measure of the estimation error in log k is the root-mean-square error, defined as

$$\varepsilon_2 = \frac{1}{M} \sum_{i=1}^{M} \left(\log k_{true,i} - \langle \log k_{est,i} \rangle \right)^{2}. \tag{B5}$$

ε_1 and ε_2 are reported in Tables 1, 2, and 3.

[87] **Acknowledgments.** The authors wish to thank Michael Cardiff and two anonymous reviewers for their constructive comments, which helped improve this paper. This material is based upon work supported as part of the Subsurface Science Scientific Focus Area funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, under award DE-AC02-05CH11231.

References

Alcolea, A., J. Carrera, and A. Medina (2006), Pilot points method incorporating prior information for solving the groundwater flow inverse problem, *Adv. Water Resour.*, *29*(11), 1678–1689, doi:10.1016/j.advwatres.2005.12.009.

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.

Allen, J. I., P. J. Somerfield, and F. J. Gilbert (2007), Quantifying uncertainty in high-resolution coupled hydrodynamic-ecosystem models, *J. Mar. Syst.*, *64*, 3–14, doi:10.1016/j.jmarsys.2006.02.010.

Anderson, R. T., et al. (2003), Stimulating the in situ activity of *Geobacter* species to remove uranium from the groundwater of a uranium-contaminated aquifer, *Appl. Environ. Microbiol.*, *69*(10), 5884–5891.

Archie, G. E. (1942), The electrical resistivity log as an aid in determining some reservoir characteristics, *Trans. Am. Inst. Min. Metall. Pet. Eng.*, *146*, 54–62.

Berre, I., M. Lien, and T. Mannseth (2009), Multi-level parameter structure identification for two-phase porous-media flow problems using flexible representations, *Adv. Water. Res.*, *32*, 1777–1788.

Caers, J. (2003), History matching under training-image based geological model constraints, *SPE J.*, *7*, 218–226.

Cardiff, M., and P. K. Kitanidis (2009), Bayesian inversion for facies detection: An extensible level set framework, *Water Resour. Res.*, *45*, W10416, doi:10.1029/2008WR007675.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and I. J. Slooten (2005), Inverse problem in hydrogeology, *Hydrogeol. J.*, *13*(1), 206–222.

Certes, C., and G. de Marsily (1991), Application of the pilot point method to the identification of aquifer transmissivities, *Adv. Water Resour.*, *14*(5), 284–300.

Cirpka, O. A., and P. K. Kitanidis (2001), Sensitivities of temporal moments calculated by the adjoint-state method and joint inversing of head and tracer data, *Adv. Water Resour.*, *24*(1), 89–103.

Dafflon, B., W. Barrash, M. Cardiff, and T. C. Johnson (2011), Hydrological parameter estimations from a conservative tracer test with variable-density effects at the Boise Hydrogeophysical Research Site, *Water Resour. Res.*, *47*, W12513, doi:10.1029/2011WR010789.

de Marsily, G. (1978), De l'identification des systemes hydrogeologiques, PhD thesis, Univ. Pierre et Marie Curie, Paris.

de Marsily, G. H. (1986), *Quantitative Hydrogeology: Groundwater Hydrology for Engineers*, Academic, San Diego, Calif.

de Marsily, G. H., G. Lavedan, M. Boucher, and G. Fasanino (1984), Interpretation of inference tests in a well field using geostatistical techniques to fit the permeability distribution in a reservoir model, in *Geostatistics for Natural Resources Characterization, Part* 2, edited by G. Verly et al., pp. 831–849, D. Reidel Pub. Co., Dordrecht, Netherlands.

Deutsch, C. V., and A. G. Journel (1992), *GSLIB: Geostatistical Software Library and User's Guide*, Oxford Univ. Press, New York.

Doherty, J. (2003), Ground water model calibration using pilot points and regularization, *Ground Water*, *41*(2), 170–177.

Doherty, J. E., M. N. 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 pp.

Englert, A. S. S. Hubbard, K. H. Williams, L. Li, and C. I. Steefel (2009), Feedbacks between hydrological heterogeneity and bioremediation induced bigeochemical transformations, *Environ. Sci. Technol.*, *43*, 5197–5204.

Fang, Y., S. B. Yabusaki, S. J. Morrison, J. E. Amonette, and P. E. Long (2009), Multicomponent reactive transport modeling of uranium bioremediation field experiments, *Geochim. Chosmochim. Acta*, *73*(20), 6029–6051.

Finsterle, S. (1999), iTOUGH2 user's guide, *Rep. LBNL-40040*, Lawrence Berkeley Natl. Lab., Berkeley, Calif.

Finsterle, S. (2004), Multiphase inverse modeling: Review and iTOUGH2 applications, *Vadose Zone J.*, *3*, 747–762.

Finsterle, S., and M. B. Kowalsky (2008), Joint hydrological-geophysical inversion for soil structure identification, *Vadose Zone J.*, *7*, 287–293, doi:10.2136/vzj2006.0078.

Gómez-Hérnandez, J. J., A. Sahuquillo, and J. E. Capilla (1997), Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data: 1. Theory, *J. Hydrol.*, *204*(1–4), 162–174.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martineza (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modeling, *J. Hydrol.*, *377*, 80–91.

Hendricks Franssen, H. J., A. Alcolea, M. Riva, M. Bakr, N. van de Wiel, F. Stauffer, and A. Guadagnini (2009), A comparison of seven methods for the inverse modelling of groundwater flow: Application to the characterisation of well catchments, *Adv. Water Resour.*, *32*(6), 851–872, doi:10.1016/j.advwatres.2009.02.011.

Hu, L. Y. (2000), Gradual deformation and iterative calibration of Gaussian-related stochastic models, *Math. Geol.*, *32*(1), 87–108.

Hu, B. X., M. M. Meerschaert, W. Barrash, D. W. Hyndman, C. M. He, X. Y. Li, and L. J. Guo (2009), Examining the influence of heterogeneous porosity fields on conservative solute transport, *J. Contam. Hydrol.*, *108*(3–4), 77–88.

Hyndman, D. W., and S. M. Gorelick (1996), Estimating lithologic and transport properties in three dimensions using seismic and tracer data: The Kesterson aquifer, *Water Resour. Res.*, *32*(9), 2659–2670.

Hyndman, D. W., J. M. Harris, and S. M. Gorelick (1994), Coupled seismic and tracer test inversion for aquifer property characterization, *Water Resour. Res.*, *30*(7), 1965–1977.

Jung, Y., R. S. Ranjithan, and G. Mahinthakumar (2011), Subsurface characterization using a D-optimality based pilot point method, *J. Hydroinf.*, *13*(4), 775–793, doi:10.2166/hydro.2010.111.

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

Kowalsky, M. B., S. Finsterle, and Y. Rubin (2004), Estimating flow parameter distributions using ground-penetrating radar and hydrological measurements during transient flow in the vadose zone, *Adv. Water Resour.*, *27*(6), 583–599.

Kowalsky, M. B., S. Finsterle, J. Peterson, S. Hubbard, Y. Rubin, E. Majer, A. Ward, and G. Gee (2005), Estimation of field-scale soil hydraulic and dielectric parameters through joint inversion of GPR and hydrological data, *Water Resour. Res.*, *41*, W11425, doi:10.1029/2005WR004237.

Kowalsky, M. B., E. Gasperikova, S. Finsterle, D. Watson, G. Baker, and S. S. Hubbard (2011), Coupled modeling of hydrogeochemical and electrical resistivity data for exploring the impact of recharge on subsurface contamination, *Water Resour. Res.*, *47*, W02509, doi:10.1029/2009WR008947.

Lavenue, A. M., and J. F. Pickens (1992), Application of a coupled adjoint sensitivity and kriging approach to calibrate a groundwater flow model, *Water Resour. Res.*, *28*(6), 1543–1569.

Lehikoinen, A., S. Finsterle, A. Voutilainen, L. M. Heikkinen, M. Vauhkonen, and J. P. Kaipio (2007), Approximation errors and truncation of computational domains with application to geophysical tomography, *Inverse Probl. Imaging*, *1*(2), 371–389.

Levenberg, K. (1944), A method for the solution of certain nonlinear problems in least squares, *Q. Appl. Math.*, *2*, 164–168.

Li, L., C. I. Steefel, M. B. Kowalsky, A. Englert, and S. S. Hubbard (2010), Effects of physical and geochemical heterogeneities on mineral transformation and biomass accumulation during biostimulation experiments at Rifle, Colorado, *J. Contam. Hydrol.*, *112*, 45–63.

Li, L., N. Gawande, M. B. Kowalsky, C. Steefel, and S. Hubbard (2011), Physicochemical heterogeneity controls on uranium bioreduction rates at the field scale, *Environ. Sci. Technol.*, *45*(23), 9959–9966, doi:10.1021/es201111y.

Linde, N., S. Finsterle, and S. Hubbard (2006), Inversion of tracer test data using tomographic constraints, *Water Resour. Res.*, *42*, W04410, doi:10.1029/2004WR003806.

Maréchal, D. (2004) A soil-based approach to rainfall-runoff modeling in ungauged catchments for England and Wales, PhD thesis, Cranfield Univ., Cranfield, U. K.

Marquardt, D. W. (1963), An algorithm for least squares estimation of nonlinear parameters, *SIAM J. Appl. Math.*, *11*, 431–441.

McLaughlin, D., and L. R. Townley (1996), A reassessment of the groundwater inverse problem, *Water Resour. Res.*, *32*(5), 1131–1161.

Moore, C., and J. Doherty (2006), The cost of uniqueness in groundwater model calibration, *Adv. Water Resour.*, *29*(4), 605–623.

Pollock, D., and O. A. Cirpka (2008), Temporal moments in geoelectrical monitoring of salt tracer experiments, *Water Resour. Res.*, *44*, W12416, doi:10.1029/2008WR007014.

Pollock, D., and O. A. Cirpka (2010), Fully coupled hydrogeophysical inversion of synthetic salt tracer experiments, *Water Resour. Res.*, *46*, W07501, doi:10.1029/2009WR008575.

Pollock, D., and O. A. Cirpka (2012), Fully coupled hydrogeophysical inversion of a laboratory salt tracer experiment monitored by electrical resistivity tomography, *Water Resour. Res.*, *48*, W01505, doi:10.1029/2011WR010779.

Pruess, K., C. Oldenburg, and G. Moridis (1999), TOUGH2 user's guide, version 2.0, *Rep. LBNL-43134*, Lawrence Berkeley Natl. Lab., Berkeley, Calif.

RamaRao, B. S., G. de Marsily, and M. G. Marietta (1995), Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. Theory and computational experiments, *Water Resour. Res.*, *31*(3), 475–493.

Rubin, Y. (2003), *Applied Stochastic Hydrogeology*, 391 pp., Oxford Univ. Press, Oxford, U. K.

Rubin, Y., X. Chen, H. Murakami, and M. Hahn (2010), A Bayesian approach for inverse modeling, data assimilation, and conditional simulation of spatial random fields, *Water Resour. Res.*, *46*, W10523, doi:10.1029/2009WR008799.

Sun, N.-Z., S.-L. Yang, and W. W.-G. Yeh (1998), A proposed stepwise regression method for model structure identification, *Water Resour. Res.*, *34*(10), 2561–2572.

Tsai, F. T.-C. (2006), Enhancing random heterogeneity representation by mixing the kriging method with the zonation structure, *Water Resour. Res.*, *42*, W08428, doi:10.1029/2005WR004111.

Tsai, F. T.-C., N.-Z. Sun, and W. W.-G. Yeh (2003), A combinatorial optimization scheme for parameter structure identification in ground water modeling, *Ground Water*, *41*(2), 156–169.

Williams, K. H., et al. (2011), Acetate availability and its influence on sustainable bioremediation of uranium-contaminated groundwater, *Geomicrobiol. J.*, *28*(5–6), 519–539, doi:10.1080/01490451.2010.520074:2011.

Wu, C.-M., T.-C. J. Yeh, J. Zhu, T. H. Lee, N.-S. Hsu, C.-H. Chen, and A. F. Sancho (2005), Traditional analysis of aquifer tests: Comparing apples to oranges?, *Water Resour. Res.*, *41*, W09402, doi:10.1029/2004WR003717.

Yabusaki, S. B., et al. (2007), Uranium removal from groundwater via in situ biostimulation: Field-scale modeling of transport and biological processes, *J. Contam. Hydrol.*, *93*, 216–235.

Yabusaki, S. B., Y. Fang, K. H. Williams, C. J. Murray, A. L. Ward, R. D. Dayvault, S. R. Waichler, D. R. Newcomer, F. A. Spane, and P. E. Long (2011), Variably saturated flow and multicomponent biogeochemical reactive transport modeling of a uranium bioremediation field experiment, *J. Contam. Hydrol.*, *126*, 271–290, doi:10.1016/j.jconhyd.2011.09.002.

Yeh, W. W.-G. (1986), Review of parameter identification procedures in groundwater hydrology: The inverse problem, *Water Resour. Res.*, *22*(2), 95–108.