AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2016WR019185

Key Points:
- Sequential kriging incorporates site-specific large-scale geologic features
- Residual covariance includes uncertainties of the geologic zones, zone properties, and heterogeneity within the zone
- Residual covariance reflecting geologic information improves hydraulic tomographic survey analysis

Supporting Information:
- Supporting Information S1

Correspondence to:
J.-C. Wen,
wenjc@yuntech.edu.tw

Citation:
Zha, Y., T.-C. J. Yeh, W. A. Illman, H. Onoe, C. M. W. Mok, J.-C. Wen, S.-Y. Huang, and W. Wang (2017), Incorporating geologic information into hydraulic tomography: A general framework based on geostatistical approach, *Water Resour. Res.*, *53*, 2850–2876, doi:10.1002/2016WR019185.

Received 9 MAY 2016
Accepted 4 MAR 2017
Accepted article online 9 MAR 2017
Published online 8 APR 2017

Incorporating geologic information into hydraulic tomography: A general framework based on geostatistical approach

Yuanyuan Zha[1,2], Tian-Chyi J. Yeh[2], Walter A. Illman[3], Hironori Onoe[4], Chin Man W. Mok[5], Jet-Chau Wen[6,7], Shao-Yang Huang[8], and Wenke Wang[9]

[1]State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, China, [2]Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, Arizona, USA, [3]Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Ontario, Canada, [4]Japan Atomic Energy Agency, Mizunami, Japan, [5]GSI Environmental Inc., Oakland, California, USA, [6]Department of Safety, Health and Environmental Engineering, National Yunlin University of Science and Technology, Douliu, Taiwan, [7]Research Center for Soil and Water Resources and Natural Disaster Prevention, National Yunlin University of Science and Technology, Douliu, Taiwan, [8]Graduate School of Engineering Science and Technology, National Yunlin University of Science and Technology, Douliu, Taiwan, [9]Key Laboratory of Subsurface Hydrology and Ecological Effects in Arid Region, Chang'an University, Xi'an, China

Abstract Hydraulic tomography (HT) has become a mature aquifer test technology over the last two decades. It collects nonredundant information of aquifer heterogeneity by sequentially stressing the aquifer at different wells and collecting aquifer responses at other wells during each stress. The collected information is then interpreted by inverse models. Among these models, the geostatistical approaches, built upon the Bayesian framework, first conceptualize hydraulic properties to be estimated as random fields, which are characterized by means and covariance functions. They then use the spatial statistics as prior information with the aquifer response data to estimate the spatial distribution of the hydraulic properties at a site. Since the spatial statistics describe the generic spatial structures of the geologic media at the site rather than site-specific ones (e.g., known spatial distributions of facies, faults, or paleochannels), the estimates are often not optimal. To improve the estimates, we introduce a general statistical framework, which allows the inclusion of site-specific spatial patterns of geologic features. Subsequently, we test this approach with synthetic numerical experiments. Results show that this approach, using conditional mean and covariance that reflect site-specific large-scale geologic features, indeed improves the HT estimates. Afterward, this approach is applied to HT surveys at a kilometer-scale-fractured granite field site with a distinct fault zone. We find that by including fault information from outcrops and boreholes for HT analysis, the estimated hydraulic properties are improved. The improved estimates subsequently lead to better prediction of flow during a different pumping test at the site.

1. Introduction

Multiscale geologic heterogeneity is the rule rather than the exception. Both small-scale heterogeneity and large-scale geologic structures or features (i.e., lithology or geologic strata, faults, folds, and lineations) can significantly impact the flow field and produce anomalous hydraulic responses at the local scale [e.g., *Ronayne et al.*, 2008].

It is well-known that geologic maps portray the spatial distribution of large-scale geologic features, and geophysical surveys can detect subsurface structures or anomalies [*Soueid Ahmed et al.*, 2015]. Nevertheless, information about such large-scale features often involves great uncertainty. For example, *Brauchler et al.* [2007] showed that geophysical structures detected by seismic tomography do not agree well with hydrogeologic structures identified by hydraulic tomography (HT). Likewise, geologic or geophysical logs are frequently used for the construction of 3-D geologic conceptual models. These models are equally ambiguous because they are built upon interpolation and extrapolation of local data from sparse boreholes [*Poeter and McKenna*, 1995]. In addition, exact hydraulic property values of these large-scale geologic features are seldom known. Consequently, our knowledge of large-scale geologic features is regarded as qualitative data in hydrogeologic modeling efforts [*Carrera et al.*, 2005; *Grana et al.*, 2012].

In spite of their qualitative nature, groundwater flow forward models often treat these large-scale geologic structures (such as facies) as well-defined zones, and some “typical” hydraulic property values are assigned to each zone according to its rock or sediment type. Equally, inverse models admit these “well-defined” geometries of the zones and then estimate the values of the hydraulic properties for each zone that reproduce observed aquifer responses [*Hendricks Franssen et al*., 2009]. In order to address the uncertainty of the geometries, *Eppstein and Dougherty* [1996] suggested that the impacts of uncertainty in zone geometry can be reduced by estimating the zonal properties and zonation simultaneously. While these approaches are practical, they often yield unsatisfactory predictions of flow and solute transport [e.g., *Konikow and Bredehoeft*, 1992; *Meier et al*., 2001; *Ni et al*., 2009; *Huang et al*., 2011].

Stochastic modeling approaches, on the other hand, adopt the stochastic conceptualization of heterogeneity [*Yeh et al*., 2015a, 2015b]. They consider the hydraulic parameter of a field site as a random field, characterized by a uniform mean and a covariance model. The mean represents the most likely values for the properties of the entire field site, while covariance represents their likely deviations, and average dimensions of these facies or zonal features. This general statistical description of a field site is then tailored to a site-specific one by employing mathematical algorithms (such as kriging or inverse models) to condition the random field such that it honors observed hydraulic properties or aquifer responses at observation locations. However, these stochastic approaches cannot fully recover the large-scale features with only hydraulic observations from sparse wells [*Meier et al*., 2001]. As a consequence, geologic features (e.g., facies, zones, faults, and karst conduits), which exhibit abrupt jumps of the hydraulic properties, are often smeared and distorted in hydrogeologic inverse results [*Tsai*, 2006; *Cardiff and Kitanidis*, 2009].

Recently developed HT surveys (see reviews by *Illman* [2014], *Cardiff et al*. [2012], and *Yeh et al*. [2015a]), which can provide high-resolution maps of the aquifer heterogeneity with a limited number of wells, also suffer from these difficulties. Over the past years, many researchers have attempted to modify the initial starting hydraulic parameter values and the regularization term in the inverse model to incorporate the knowledge of large-scale geologic structures. For example, *Soueid Ahmed et al*. [2015] proposed the use of structural information inferred from a guiding image to improve the estimates of hydraulic conductivity in steady state HT. The guiding image can be obtained from the geophysical or geologic survey. The extracted structural information is introduced as a weighted four-direction smoothing for regularization. The synthetic experiments show that image-guided inversions are more accurate and yield results at a higher resolution than those based on classical regularization models. The illustrative examples are 2-D cases, and simple heterogeneity with several zones (each zone assigned with one conductivity value) are considered. Although the impacts of the inaccurate structure information or inaccurate mean are tested in the examples, quantitative analysis of these uncertain factors is omitted.

Fienen et al. [2008] proposed an interactive Bayesian geostatistical inverse protocol for hydraulic tomography based on quasilinear geostatistical approach (QLGA) [*Kitanidis*, 1995]. They show that explicit trade-off between data misfit and the prior assumption is controlled by the ratio of the parameter variance and the epistemic error, which is similar to the concept of dynamic stabilizer used in the Levenberg-Marquardt algorithm [*Pujol*, 2007]. In order to handle discontinuities in the parameter field, *Fienen et al*. [2008] proposed an interactive algorithm to select the thresholds for the zonation. Different zones with significantly different mean values were characterized by the same prior covariance function, and this single covariance function was modified so that two locations at different zones had zero correlation. Numerical tests revealed that with zonation assumptions, the inversion obtained a better result than without zonation, which suffered from the smoothing of the zone boundaries. However, the modified single covariance function based on the assumption that all zones have the same variance and correlation structure may not be adequate to characterize different variabilities in different zones of the geologic formations. Furthermore, the method omits the uncertainty of the zone boundaries.

Cardiff and Kitanidis [2009] proposed a method for defining facies locations and boundaries using the level set method and for moving the boundaries between zones using a gradient-based technique that improves fitting through iterative deformation of the boundaries. The level set representation of facies boundaries is flexible to deal with geologic facies in any shape, size, or number. While the proposed method is general in handling facies-dominated problems, the parameter variability inside each facies was not considered. Also, the variability of covariance function between different zones was not discussed.

Tso et al. [2016] investigated the relative importance of prior information and different types of hydraulic data (flux or drawdown) in HT survey. They concluded that when large-scale trend information is available, the distributed mean and a simple covariance function using short correlation scales yield the best results. The approach proposed by *Tso et al.* [2016] was subsequently validated by sandbox and field experiments [*Zhao et al.*, 2016; *Zhao and Illman*, 2017], and they found that the value of geologic information is most useful when the numbers of pumping tests and monitoring points are sparse. The influence of unknown zone boundary was also investigated in *Tso et al.* [2016] by smearing the true boundary with a sinusoidal wave. However, using a simple covariance model with short correlation scale may not be suitable to describe the situation when the zone boundary is not known exactly and the small-scale heterogeneity in each zone differs from the others.

Previous works have highlighted the benefits of incorporating site-specific geologic structure information into groundwater models when HT data are limited. However, quantitative incorporation of the site-specific information by using more general mean and covariance function remains a challenge. Specifically, site-specific geologic information often includes both structural geometries (i.e., the shape of facies and zones, etc.) and their associated hydraulic properties (e.g., high hydraulic conductivity (K) and low K values). A general approach to incorporate these two types of information into classical geostatistical approaches has not been fully resolved. Furthermore, previous studies mainly focused on synthetic aquifers. Quantitative evaluation of the impact of incorporating geologic information in three-dimensional field problems still needs to be explored.

This paper aims to present a general geostatistical framework to address the aforementioned issues. It is organized as follows: in section 2, we review the Successive Linear Estimator (SLE) developed by *Yeh et al.* [1996], *Zhang and Yeh* [1997], *Hughson and Yeh* [2000], *Yeh and Liu* [2000], and *Zhu and Yeh* [2005], which is one of the algorithms for HT analysis developed in a Bayesian perspective (section 2.1). We then introduce a general method to derive conditional mean and conditional covariance, which account for site-specific geologic information (section 2.2). These mean and covariance can be used in HT analysis as prior information. In section 3, we design two synthetic examples to demonstrate the proposed method and to investigate the influence of incorporating site-specific geologic information on HT result. Section 4 presents a field HT application of the proposed method, followed by conclusions in section 5.

2. Methods

A method of inverse modeling is required to identify parameter fields, such as the hydraulic conductivity distribution, given a series of hydraulic responses collected during a HT survey. Derived in the Bayesian framework, QLGA [*Kitanidis*, 1995; *Fienen et al.*, 2009] and SLE [*Yeh et al.*, 1996] are widely employed to interpret HT data. Since groundwater flow inverse problems are in general nonlinear, iterative procedures are involved in these two methods. SLE conceptualizes the iterative procedure as a sequence of Bayesian updating steps and update both the mean and covariance of the parameter fields during iteration. In contrast, QLGA successively linearizes the forward problem about the current estimate, while keeping the covariances [*Nowak and Cirpka*, 2004]. Despite the differences, the two widely used approaches share many common features [*Liu et al.*, 2014; *Soueid Ahmed et al.*, 2015]. For instance, they both require the prior mean and covariance of the parameter field, and they provide the uncertainty estimation besides the parameter distribution. For completeness, we first review the SLE algorithm. Afterward, a general framework to construct the prior mean and covariance (required inputs in SLE) based on geologic information is proposed. A flowchart that summarizes the framework is provided in Figure 1.

2.1. Successive Linear Estimator (SLE)

Stochastic inverse models including SLE adopt a highly parameterized heterogeneous conceptual model, which usually discretizes the 3-D domain of the field site into N elements. The hydraulic parameter of the N elements (e.g., natural logarithm of K, ln$K$) is composed of a $(N \times 1)$ vector. The model then considers these hydraulic parameters as spatial stochastic processes (random fields) with prior (unconditional) mean $\mathbf{Y}$ $(N \times 1)$, and the prior perturbations $\mathbf{y}$ $(N \times 1)$, respectively. These perturbations represent the spatial variability of the parameters.

SLE estimates the most likely parameter value (i.e., conditional effective value) for each element, given (conditioned with) the observed drawdown (or head) data from the HT survey. Suppose during an HT survey,

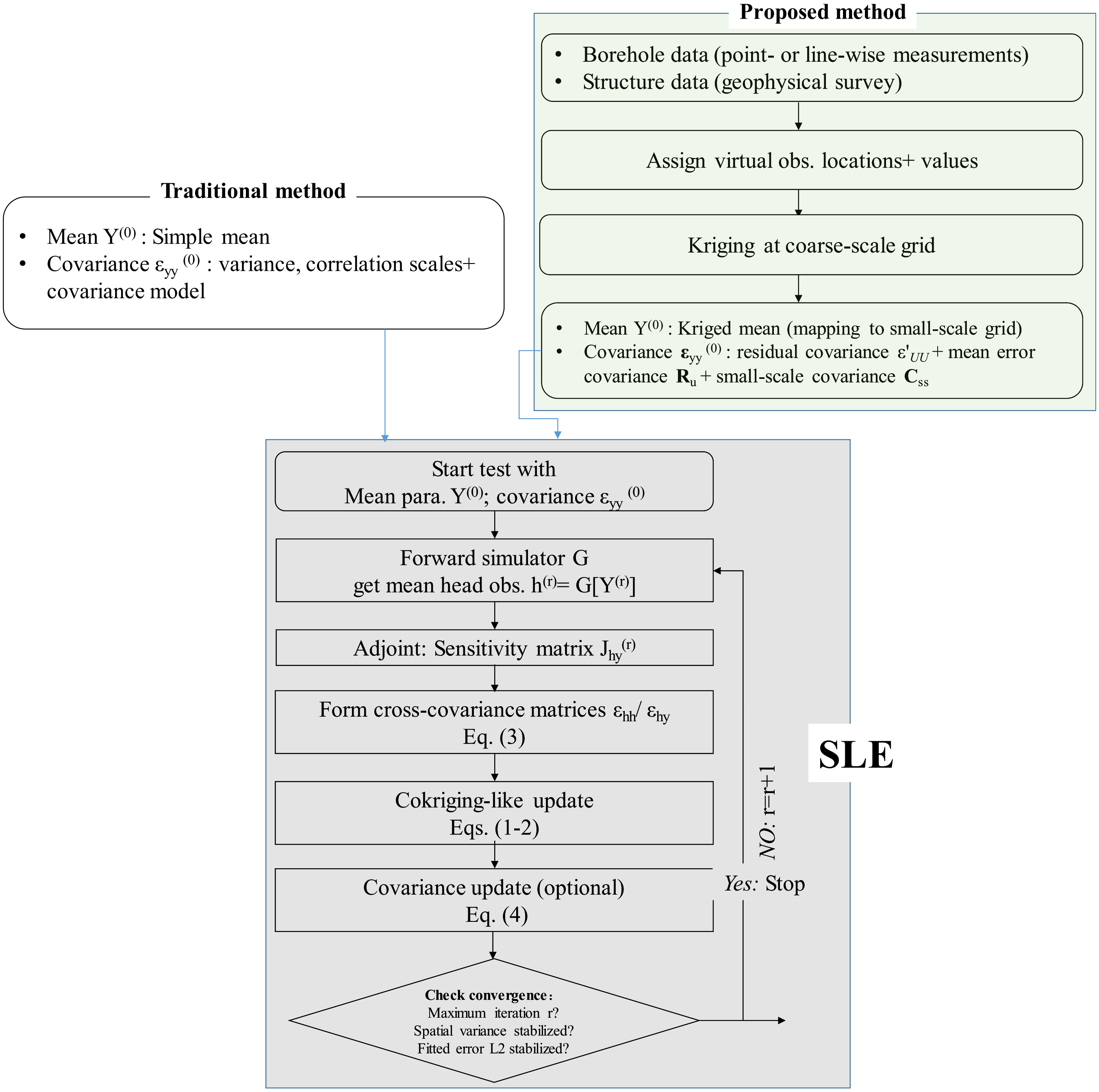

Figure 1. Flowchart summarizing the SLE inverse algorithm and the proposed method that provding the prior mean and covariance function for SLE.

we have collected M observed heads in space and time, denoted by the data vector $\mathbf{d}$. The estimates of parameter fields, given the observation, is $\hat{\mathbf{Y}}_c$ (subscript c denotes conditional), are iteratively determined using the following linear estimator [*Yeh et al.*, 1996]:

$$\hat{\mathbf{Y}}_c^{(r+1)}=\hat{\mathbf{Y}}_c^{(r)}+\boldsymbol{\omega}^{\mathrm{T}}\left(\mathbf{d}-\mathbf{G}\left(\hat{\mathbf{Y}}_c^{(r)}\right)\right) \tag{1}$$

where r is the iteration index; $\mathbf{G}(\cdot)$ indicates the nonlinear relationship between $\mathbf{Y}$ and $\mathbf{d}$ (i.e., a forward groundwater flow model), which produces the simulated heads at the observation locations and times using the parameters obtained at iteration r. The coefficient matrix, $\boldsymbol{\omega}$ ($M \times N$), denotes the weights, which assigns the contribution of the difference between the observed and simulated head at each observation location and time to previously estimated parameter values at each element. The superscript T denotes the transpose.

The coefficient matrix $\boldsymbol{\omega}$ is determined by solving the following equation [*Yeh et al*., 1996]:

$$\left[\boldsymbol{\varepsilon}_{dd}^{(r)}+\mathbf{R}\right]\boldsymbol{\omega}^{(r)}=\boldsymbol{\varepsilon}_{dy}^{(r)} \tag{2}$$

where $\mathbf{R}$ is the covariance matrix ($M \times M$) of the measurement error associated with head measurements. The solution of equation (2) requires the knowledge of covariance $\boldsymbol{\varepsilon}_{dd}$ and cross-covariance $\boldsymbol{\varepsilon}_{dy}$, which can be derived from the first-order numerical approximation [*Yeh et al*., 1996]:

$$\begin{aligned}\boldsymbol{\varepsilon}_{dd}^{(r)}&=\mathbf{J}_d^{(r)}\boldsymbol{\varepsilon}_{yy}^{(r)}\mathbf{J}_d^{(r)\mathrm{T}},\\ \boldsymbol{\varepsilon}_{dy}^{(r)}&=\mathbf{J}_d^{(r)}\boldsymbol{\varepsilon}_{yy}^{(r)}\end{aligned} \tag{3}$$

where $\mathbf{J}_\mathrm{d}$ ($M \times N$) is the sensitivity (Jacobian) matrix of head data with respect to the element-wise parameter using the parameters estimated at the current iteration. At the beginning of iteration (when $r=0$), $\boldsymbol{\varepsilon}_{yy}$ is the unconditional covariance matrix of parameters $\mathbf{Y}$, which is traditionally constructed by a prior variance, correlation lengths, and a covariance model (see section 2.1.3). After that ($r\geq 1$), the residual or conditional covariance function of parameters is updated as [*Yeh and Liu*, 2000]:

$$\boldsymbol{\varepsilon}_{yy}^{(r+1)}=\boldsymbol{\varepsilon}_{yy}^{(r)}-\boldsymbol{\omega}^{(r)\mathrm{T}}\boldsymbol{\varepsilon}_{dy}^{(r)} \tag{4}$$

The SLE bears the concept of cokriging or stochastic linear estimator equation (e.g., unbiased estimates with minimum variance). The nonlinearity between parameters and heads is dealt with successive iteration. At iteration $r=0$, SLE requires guessed values for the mean ($\hat{\mathbf{Y}}_c^{(0)}$ used in eqaution (1)) and covariance function ($\boldsymbol{\varepsilon}_{yy}^{(0)}$ used in equations (3) and (4)). In the view of Bayesian statistics, $\hat{\mathbf{Y}}_c^{(0)}=\mathbf{Y}_\mathrm{prior}$ and $\boldsymbol{\varepsilon}_{yy}^{(0)}=\mathbf{C}_\mathrm{prior}$ are the prior information of the unknown parameter field. Afterward, SLE updates the mean and the covariance at each iteration due to the gradual assimilation of the observation information, and reduces the uncertainty of the estimate.

2.1.1. Maximum A Posteriori Approach

SLE is similar to the maximum a posteriori (MAP) inverse approach, but it is different. As pointed out by *Carrera and Glorioso* [1991], the cokriging equation produces the same estimate of the first iteration of maximum posterior approaches, if the initial guess mean is taken as the prior. Let $p(\mathbf{Y}|\mathbf{d})$ be the probability density function (pdf) of model parameter $\mathbf{Y}$ conditioned on the data set $\mathbf{d}$ and $p(\mathbf{Y})$ is the prior pdf. Bayes theorem gives the pdf of model parameter $\mathbf{Y}$ after the assimilation of the data $\mathbf{d}$ [*Fienen et al*., 2009]:

$$p(\mathbf{Y}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{Y})p(\mathbf{Y}) \tag{5}$$

If the prior pdf $p(\mathbf{Y})$ can be approximated as Gaussian with mean $\mathbf{Y}_\mathrm{prior}$ ($N \times 1$) and covariance $\mathbf{C}_\mathrm{prior}$ ($N \times N$), and the error in observation $\mathbf{d}$ ($M \times 1$) are normally distributed with zero mean and covariance $\mathbf{R}$ ($M \times M$), equation (5) becomes

$$\ln p(\mathbf{Y}|\mathbf{d}) \propto -\frac{1}{2}\left[(\mathbf{G}(\mathbf{Y})-\mathbf{d})^T\mathbf{R}^{-1}(\mathbf{G}(\mathbf{Y})-\mathbf{d})+\left(\mathbf{Y}-\mathbf{Y}_\mathrm{prior}\right)^T\mathbf{C}_\mathrm{prior}^{-1}\left(\mathbf{Y}-\mathbf{Y}_\mathrm{prior}\right)\right] \tag{6}$$

Using the Gauss-Newton method to minimize the objective function –ln $p(\mathbf{Y}|\mathbf{d})$, the $(r+1)$th iterative estimate of parameter $\mathbf{Y}$ is [*Chen and Oliver*, 2013]:

$$\hat{\mathbf{Y}}_c^{(r+1)}=\hat{\mathbf{Y}}_c^{(r)}+\left(\mathbf{J}_d^{(r)\mathrm{T}}\mathbf{R}^{-1}\mathbf{J}_d^{(r)}+\mathbf{C}_\mathrm{prior}^{-1}\right)^{-1}\left[\mathbf{J}_d^{(r)\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{d}-\mathbf{G}\left(\hat{\mathbf{Y}}_c^{(r)}\right)\right)+\mathbf{C}_\mathrm{prior}^{-1}\left(\mathbf{Y}_\mathrm{prior}-\hat{\mathbf{Y}}_c^{(r)}\right)\right] \tag{7}$$

with estimation covariance of

$$\boldsymbol{\varepsilon}_{yy}^{(r+1)}=\left(\mathbf{J}_d^{(r)\mathrm{T}}\mathbf{R}^{-1}\mathbf{J}_d^{(r)}+\mathbf{C}_\mathrm{prior}^{-1}\right)^{-1} \tag{8}$$

Through several linear algebraic manipulations (see equations ((1).106) and (1.107) in *Tarantola* [2005]), we have

$$\hat{\mathbf{Y}}_c^{(r+1)}=\hat{\mathbf{Y}}_c^{(r)}+\left(\mathbf{Y}_{\text{prior}}-\hat{\mathbf{Y}}_c^{(r)}\right)+\mathbf{C}_{\text{prior}}\mathbf{J}_d^{(r)\mathrm{T}}\left(\mathbf{J}_d^{(r)}\mathbf{C}_{\text{prior}}\mathbf{J}_d^{(r)\mathrm{T}}+\mathbf{R}\right)^{-1}\left[(\mathbf{d}-\mathbf{G}(\mathbf{Y}))+\mathbf{J}_d^{(r)}\left(\mathbf{Y}_{\text{prior}}-\hat{\mathbf{Y}}_c^{(r)}\right)\right] \tag{9}$$

$$\boldsymbol{\varepsilon}_{yy}^{(r+1)}=\mathbf{C}_{\text{prior}}-\mathbf{C}_{\text{prior}}\mathbf{J}_d^{(r)\mathrm{T}}\left(\mathbf{J}_d^{(r)}\mathbf{C}_{\text{prior}}\mathbf{J}_d^{(r)\mathrm{T}}+\mathbf{R}\right)^{-1}\mathbf{J}_d^{(r)}\mathbf{C}_{\text{prior}} \tag{10}$$

By comparing equations (1–4) and equations (9) and (10), it is found that the calculated $\hat{\mathbf{Y}}_c^{(r+1)}$ and $\boldsymbol{\varepsilon}_{yy}^{(r+1)}$ have the same forms in SLE and MAP formulations if we set initial guess mean $\hat{\mathbf{Y}}_c^{(0)}$ as the prior $\mathbf{Y}_{\text{prior}}$. In other words, the first iteration of SLE and MAP yield the same estimated mean and covariance. However, after the first iteration, SLE uses the updated mean (e.g., substituting $\mathbf{Y}_{\text{prior}}$ by $\hat{\mathbf{Y}}_c^{(r)}$ in the equation (9)) and updated covariance function (e.g., substituting $\mathbf{C}_{\text{prior}}$ by $\boldsymbol{\varepsilon}_{yy}^{(r)}$ in the equation (10)) as prior information. That is, MAP uses static prior information, while SLE recursively updates the mean and covariance from the last iteration as the prior (This is similar to Kalman filter in signal analysis, where new observation in time is added). In other words, a posterior mean and covariance at iteration r serve as a prior at iteration $r+1$. The logic behind this is that the inverse model gradually learns from observation data for every iteration, and this changes the pdf of the uncertainty associated with the estimates at every iteration [*Yeh et al*., 1996; *Yeh and Liu*, 2000].

2.1.2. Stabilizer Term

For solving the nonlinear problem, a stabilizer term is usually added to ensure the stability of equation (2):

$$\left[\boldsymbol{\varepsilon}_{dd}^{(r)}+\mathbf{R}+\theta^{(r)}\text{diag}(\boldsymbol{\varepsilon}_{dd}^{(r)})\right]\boldsymbol{\omega}^{(r)}=\boldsymbol{\varepsilon}_{dy}^{(r)} \tag{11}$$

The θ is a dynamic stability multiplier, and diag(ε_{dd}) is a diagonal matrix with the same diagonal elements as ε_{dd} [*Dietrich and Newsam*, 1989; *Carrera and Glorioso*, 1991; *Yeh and Liu*, 2000]. Mathematically, the dynamic stabilizer term facilitates the solution switching between the Gauss-Newton solution and the steepest-descent method, which is known as the Levenberg-Marquardt approach [*Pujol*, 2007]. The reason is that when the initial guess values are far away from the true solutions for a nonlinear problem, the Gauss-Newton solution (without the stabilizer term) may lead to the divergence of the solution if a full step is taken to update the parameters. From a Bayesian perspective, the stabilizer term is similar to the concept of epistemic error for imperfect observations [e.g., *Fienen et al*., 2008] and conceptual model error (e.g., inaccurate boundary conditions, inaccurate sensitivity matrix evaluated at the estimated mean $\hat{\mathbf{Y}}_c^{(r)}$, which may be far away from true values). In order to prevent the model from diverging due to these unpredictable errors from observation data or conceptual models, the stabilizer is used to dampen the change of the prior parameter field. Since initially the guessed $\mathbf{Y}$ is far away from true solution and tends to produce larger conceptual model errors, generally a larger θ value is preferred initially, and then the value decreases with iteration. Note that since $\mathbf{R}$ is usually unknown, it is often dropped in equation (11) and the stabilizer implicitly includes the data error. However, if a rough magnitude of error is known, we can control the iteration process by either setting a minimal stabilizer term or by manually checking the misfit of data to avoid overfitting problem [*Xiang et al*., 2009]. The misfit of data can be quantitatively evaluated by different norms. Two criteria, the average absolute (L1) and the mean squared error (L2) norms are often used to evaluate the differences between the observed and simulated heads:

$$L1=\frac{1}{n}\sum_{i=1}^{n}|h_i-h'_i| \tag{12}$$

$$L2=\frac{1}{n}\sum_{i=1}^{n}(h_i-h'_i)^2 \tag{13}$$

where h_i and h'_i represents the observed head and simulated head at location or time i, and n is the total number of observations.

2.1.3. Prior Mean and Correlation Function

As discussed above, SLE and other Bayesian geostatistical methods require the knowledge of prior mean and covariance function. The importance of the prior information is self-evident in Bayes theorem: if the data is not sufficient to reproduce the true field, both prior mean and covariance serve as additional constraints and the choice of the prior matters.

Typically, a uniform (unconditional) mean value and unconditional covariance functions of a geologic model are used for $\mathbf{Y}_{\text{prior}}$ in a geostatistical estimation approach (such as kriging, cokriging, SLE, or QLGA). The word "unconditional" implies that HT data have not been included yet, and only general information (such as soil/rock types and average sizes of the blocks) is available based on a first glance of the site [*Yeh et al.*, 2015b]. This general prior information is then conceptualized by a general stochastic description of spatial variability of parameters. For example, a single mean value and an exponential covariance function cov(σ^2, $\mathbf{L}$) can be used to describe the prior information:

$$Y_{\text{prior}}(\mathbf{x}=\mathbf{x}_i)=m$$
$$C_{\text{prior}}(\mathbf{x}_i,\mathbf{x}_j)=\text{cov}\left(\sigma^2,\mathbf{L}\right)=\sigma^2\exp\left(-\sqrt{\frac{\Delta x^2}{L_x^2}+\frac{\Delta y^2}{L_y^2}+\frac{\Delta z^2}{L_z^2}}\right) \tag{14}$$

where m is the mean, σ^2 is the variance, and $\mathbf{L}=[L_x, L_y, L_z]^{\text{T}}$ are the correlation scales in different directions, $\Delta\mathbf{x}=\mathbf{x}_i-\mathbf{x}_j$ is the vector defined by two points $\mathbf{x}_i$ and $\mathbf{x}_j$ in 3-D space. This is equivalent to state that before any inverse modeling effort, we know the mean, variance, and general spatial structure of the geologic medium. These are merely additional constraints in a statistical sense to the inverse problem [see *Yeh et al.*, 2015b].

Note that the above unconditional mean and covariance represent the spatial statistics of all possible realizations of heterogeneous property fields, which include variability at all scales (i.e., those with or without trends, nonstationary, or stationary fields) in a geologic medium [*Yeh et al.*, 2015b]. The question then becomes how to define the spatial statistics that tailor to realizations, which contain known, but imprecise discrete large-scale structures (such as different geologic facies, folds, or faults) in the parameter field. More importantly, how to incorporate them into inverse models becomes an issue as noticed by *Fienen et al.* [2008] and *Cardiff and Kitanidis* [2009]. This issue becomes more critical when only sparse head data are available. Below, we propose a general methodology for resolving this issue.

2.2. Incorporating Site-Specific Geologic Information

A general geostatistical approach based on sequential kriging is proposed to incorporate site-specific geologic information into stochastic inverse models. This approach adopts a nested covariance function concept for two-scale heterogeneity. In other words, geologic facies, sediment layers or stratigraphy, rock types, or geophysical attributes from well logs, outcrops, or other surveys are viewed as large-scale properties, which can be conceptualized as a stochastic field characterized by a mean and a large-scale covariance function, representing the generic spatial distribution of possible geologic facies in a geologic medium. The interface between different facies or zones may exhibit an abrupt change of hydraulic properties. On the other hand, the variability inside a particular facies or zone (compared to its large-scale mean) can be described by a zero mean and a small-scale covariance function.

2.2.1. Nested Covariance Function for Multiscale Heterogeneity

Suppose geologic or geophysical well logs reveal some strata of gravel, sand, silt, and clay along boreholes in a geologic medium and only ranges of their hydraulic properties are known. A sequential kriging method can be used to include this soft (qualitative) information as prior information in inverse modeling. First, we conceptualize the hydraulic parameter Y of the aquifer as a random field consisting of coarse and fine-scale variabilities:

$$Y(\mathbf{x})=U(\mathbf{x})+S(\mathbf{x}) \tag{15}$$

where U represents the random field observed at coarse scales (e.g., the average hydraulic property of layers, stratifications, or facies). It can be decomposed into a mean and perturbation: $U(\mathbf{x})=m+u(\mathbf{x})$, $E[U(\mathbf{x})]=m$, $E[u(\mathbf{x})]=0$, and it is characterized by a spatial covariance function,

$$C_{uu}\left(\mathbf{x}_i,\mathbf{x}_j\right)=\text{cov}\left(\sigma_U^2,\mathbf{L}_U\right) \tag{16}$$

where σ_U^2 is the variance of $u(\mathbf{x})$, and $\mathbf{L}_U$ represents the spatial correlation scales of the coarse-scale property. Likewise, $S(\mathbf{x})$ denotes variation of hydraulic properties at the fine scale, which is superimposed on the hydraulic property at the coarse scale. $S(\mathbf{x})$ has a zero mean: $E[S(\mathbf{x})]=0$, and a spatial covariance function:

$$C_{ss}\left(\mathbf{x}_i,\mathbf{x}_j\right)=\text{cov}\left(\sigma_S^2,\mathbf{L}_S\right) \tag{17}$$

where σ_s^2 is the variance of S and $\mathbf{L}_s$ are its spatial correlation scales, which are much smaller than $\mathbf{L}_U$. Since small-scale heterogeneity is defined inside a particular zone or layer, different covariance functions of equation (17) with different variances and correlation scales can be assigned to different zones. Theoretically, the proposed concept could be expanded to include additional scales of heterogeneity.

2.2.2. Two-Level Material Grid and Geologic Information Conditioning

In order to implement such a two-scale variability, we discretize the medium into N_c number of coarse grids, and each of the coarse grid contains many fine grids. As a result, there are N_c coarse grids and N fine grids in the domain. The parameter value of each coarse grid is the mean of different parameter values of the fine grids in the coarse grid. The variability of the coarse-grid parameter is described by its mean, m, and covariance, $\mathbf{C}_{UU}$. The variability of the parameter of fine grids is modeled by the covariance, $\mathbf{C}_{SS}$. The parameters at fine grids (or the so-called material grid) dictate the resolution of the heterogeneity in our model. Note that these material grids (to which parameter values are assigned) can be different from the numerical meshes (for calculating state variable on the node) used in the finite element flow models.

The use of the coarse grid is for implementing the large-scale geologic information. Specifically, we assume N_o coarse-scale "virtual" observations located at the centers of the coarse grids, $\mathbf{x}_i^c$, where $i=1, 2, \ldots, N_o$, and $N_o \leq N_c$ (the total number of coarse grids in the domain). The number, locations, and the parameter values of virtual observations are assigned based on our qualitative knowledge of the geologic information. Since the locations and values of the virtual observations are arbitrary based on user specification, they can carry a wide variety of geologic information, such as K estimates from hydraulic tests, descriptions of rock types from lithology, resistivity, density logs from few boreholes, sediment facies distribution from geologic maps, and field observations of fracture/fault zones or caverns in the domain. The concept and implementation procedure of the virtual observation will be demonstrated using two illustrative examples in section 3.

Next, starting from an unconditional mean m, we construct the conditional mean field at the coarse grids using kriging with the N_o virtual observations. For convenience, we define new variables, $y=Y-m$ as the perturbation at the fine grid, and $u=U-m$ as the perturbation at the coarse grid. The kriging equation (the linear stochastic estimator) is,

$$\hat{u}\left(\mathbf{x}_j^c\right)=\sum_{i=1}^{N_o}\omega_{ij}\,\tilde{u}\left(\mathbf{x}_i^c\right) \tag{18}$$

where $\tilde{u}\left(\mathbf{x}_i^c\right)$ is the virtual observation value for the coarse-grid property based on the translation of geologic or geophysical information at the observed location $\mathbf{x}_i$, and $\hat{u}\left(\mathbf{x}_j^c\right)$ is the kriging estimate at any location, $\mathbf{x}_j^c$. Accordingly, the kriged estimates $\hat{Y}$ at coarse grids will be $\hat{u}+m$. The kriging estimate is our inferred mean parameter distribution based on site-specific geologic information with the help of assigned virtual observations. The kriging coefficients in equation (18) are calculated as,

$$\sum_{j=1}^{N_o}C_{uu}\left(\mathbf{x}_i^c,\mathbf{x}_j^c\right)\omega_{ik}=C_{uu}\left(\mathbf{x}_i^c,\mathbf{x}_k^c\right) \tag{19}$$

where $i=1,2,\cdots N_0$, $k=1,2,\cdots N_c$. Afterward, the residual covariance of u is,

$$\varepsilon'_{uu}\left(\mathbf{x}_i^c,\mathbf{x}_j^c\right)=C_{uu}\left(\mathbf{x}_i^c,\mathbf{x}_j^c\right)-\sum_{k=1}^{No}\omega_{ki}C_{uu}\left(\mathbf{x}_k^c,\mathbf{x}_i^c\right) \tag{20}$$

where $i=1,2,\cdots N_c$, $j=1,2,\cdots N_c$. This residual covariance, $\boldsymbol{\varepsilon}'_{uu}$, represents possible deviations of the estimate, $\hat{u}\left(\mathbf{x}_j^c\right)$, from $u(\mathbf{x})$ due to an insufficient number of virtual observations (i.e., $N_o \leq N_c$). It can be used to describe the boundary uncertainty of zones that have different mean values. However, it should be noted that the covariances $\boldsymbol{\varepsilon}'_{uu}$ and $\mathbf{C}_{uu}$ involved in equations (20) and (16) are most suitable for spatially continuous random fields or a domain that consists of many facies, fault zones, caverns, or paleochannels where ergodicity can be met. That is, in the field with only a few large-scale distinct zones, the spatial covariances of the field (one realization) is not the same as covariances $\boldsymbol{\varepsilon}'_{uu}$ and $\mathbf{C}_{uu}$ in geostatistics, which are built upon an infinite number of realizations [*Yeh et al.*, 2015b]. Therefore, it is not surprising that the predicted mean and uncertainty (equations (18) and (20)) via kriging may not be optimal when we compare the estimate with the true field (one realization).

For a domain consisting of only a few individual discrete anomalies (site-specific discrete features), where ergodicity is not met, we propose a new covariance function $\mathbf{R}_U$ for the situation, when zone geometry is presumed. This $\mathbf{R}_U$ is a covariance ($N_c \times N_c$) denoting the deviation of the estimated parameter value, $\tilde{u}$, from the true value u over a site-specific feature (or zone). Specifically, $R_u\left(\mathbf{x}_i^c,\mathbf{x}_j^c\right)$ is given as a positive constant value if $\mathbf{x}_i^c$ and $\mathbf{x}_j^c$ are in the same zone, Ω_k and zero otherwise. Mathematically, it is

$$R_U\left(\mathbf{x}_i^c,\mathbf{x}_j^c\right)=\begin{cases} A_k & \mathbf{x}_i^c \in \Omega_k, \mathbf{x}_j^c \in \Omega_k \\ 0 & otherwise \end{cases} \tag{21}$$

where $A_k \in (0,\sigma_U^2)$ is a constant in zone k, and $k = 1, 2, \ldots, N_\Omega$, where N_Ω is the total number of presumed zones. According to equation (3), if ε_{yy} (covariance used in HT) is given as $\mathbf{R}_U$, we can conclude that $\varepsilon_{dy}(\mathbf{x}_0, \mathbf{x}_{j1}) = \varepsilon_{dy}(\mathbf{x}_0, \mathbf{x}_{j2})$ when $\mathbf{x}_{j1}$ and $\mathbf{x}_{j2}$ are in the same zone (i.e., $\mathbf{x}_{j1} \in \Omega_k$ and $\mathbf{x}_{j2} \in \Omega_k$). Equation (3) thus yields the same cross-variance value for elements in the same zone, regardless of the sensitivity. This implies the same weight in SLE (equation (1)) for locations in the same zone. In other words, this particular covariance $\mathbf{R}_U$ carries the information of zone geometry and it regularizes the inversion process to find the appropriate mean K value for each zone. The property of $\mathbf{R}_U$ is of great importance for many practical problems. For example, geologic mapping or geophysical surveys of a field site may identify one or two large-scale geologic features, which manifest as relatively high and low K zones in an area, but exact values of the hydraulic parameters of these features are not known.

The large-scale residual covariance after the inclusion of site-specific geologic information is a combination of $\boldsymbol{\varepsilon}'_{uu}$ (coarse-scale zone geometry uncertainty) and $\mathbf{R}_U$ (mean distribution uncertainty under a presumed zone geometry):

$$\varepsilon_{uu}\left(\mathbf{x}_i^c,\mathbf{x}_j^c\right)=\varepsilon'_{uu}\left(\mathbf{x}_i^c,\mathbf{x}_j^c\right)+R_u\left(\mathbf{x}_i^c,\mathbf{x}_j^c\right) \tag{22}$$

These kriged mean (equation (18)) and large-scale residual covariance function (equation (22)) at coarse grids are next mapped onto fine grids. The mean values of Y defined at elements are constructed by:

$$\hat{Y}\left(\mathbf{x}_j\right)=\hat{u}\left(\mathbf{x}_i^c\right)+m, \quad \mathbf{x}_j \in \mathbf{x}_i^c \tag{23}$$

where $\mathbf{x}_j$ without superscript c denotes the coordinates of the fine grids. The residual covariance of Y is the summation of the large-scale covariance ($\boldsymbol{\varepsilon}_{uu}$) mapping to fine grids and the small-scale covariance ($\mathbf{C}_{ss}$):

$$\varepsilon_{YY}\left(\mathbf{x}_i,\mathbf{x}_j\right)=\varepsilon_{uu}\left(\mathbf{x}_m^c,\mathbf{x}_n^c\right)+C_{ss}\left(\mathbf{x}_i,\mathbf{x}_j\right) \tag{24}$$

where $\mathbf{x}_i \in \mathbf{x}_m^c, \mathbf{x}_j \in \mathbf{x}_n^c$. As a result, the above approach considers the residual covariance $\boldsymbol{\varepsilon}_{yy}$ resulting from three different sources: within layers' fine-scale variability ($\mathbf{C}_{SS}$), error associated with the guess mean value for coarse-scale properties under a presumed zone geometry ($\mathbf{R}_U$), and finally, uncertainty due to insufficient coarse-scale information ($\boldsymbol{\varepsilon}'_{uu}$). While the general concept of the above approach bears some similarities to the work by *Fienen et al*. [2008] and others, it differs in both concept and the methodology and many other aspects. More importantly, our proposed approach is general and can be applied to different cases as illustrated in the following examples.

It is noteworthy that the virtual observation concept is different from the well-known pilot point method. For reducing computational cost and making the inverse problem well-defined [*Yeh et al*., 2015a, 2015b], the pilot point method uses only a few selected pilot points, where hydraulic parameter values are estimated by a nonlinear algorithm minimizing the simulated and observed hydraulic head differences. The entire parameter field is obtained afterward by kriging based on the parameter values at the pilot point locations and the unconditional covariance function of the parameter [*McLaughlin and Townley*, 1996; *Soueid Ahmed et al*., 2015]. That is, the final parameter field and the estimated parameters at pilot points are not linked by the governing flow equation, and the result thus could be suboptimal [*Huang et al*., 2011]. In contrast, the virtual points here are employed to incorporate large-scale geologic information, and the estimated mean and covariance function for parameters at coarse grids then serve as the prior information for HT inversion. In other words, the virtual points are only used in constructing the prior mean and covariance; they are observations, not estimated parameters as are those in the pilot point method.

2.2.3. Extreme Examples That Account for Single Uncertainty Source

To better understand different sources of uncertainty (i.e., the covariances $\boldsymbol{\varepsilon}'_{uu}$, $\mathbf{R}_u$ and $\mathbf{C}_{SS}$), we provide some extreme examples that focus on a single source of uncertainty.

1. **$\mathbf{R}_u$-dominated scenario.** In some situations, the parameter field exhibits strong contrast between different zones, and the heterogeneity within the zones is relatively minor (i.e., $\mathbf{C}_{SS}\approx\mathbf{0}$). In addition, the zone boundaries may be identified through a geophysical investigation [e.g., *Soueid Ahmed et al.*, 2015]. In this case, we place qualitative observations at every coarse grid such that the shapes of the geologic features are known ($\boldsymbol{\varepsilon}'_{uu}=\mathbf{0}$), but the values of the hydraulic parameters of each feature are unknown. This implies that one must guess the observation value, $\tilde{u}$, which likely will deviate from the true u. One extreme, but commonly considered case is that $\tilde{u}$ values are chosen to be zero, so that there is no information about the large-scale mean (U), and the constant in $\mathbf{R}_u$ is equal to the ensemble variance. This essentially leads to the zonation model [*Zhou et al.*, 2014], in which the covariance $\mathbf{R}_u$ makes sure that during parameter updating, the parameter updating increments for elements in one zone during inversion are always the same, since every two of them have a correlation of 1.
2. **$\boldsymbol{\varepsilon}'_{uu}$-dominated scenario.** This extreme scenario corresponds to the cases when partial large-scale zone structures are unknown. For instance, the geologic stratification is obtained from several well logs. The zone spatial distribution has to be interpolated and it involves uncertainty. As a result, the elements of $\boldsymbol{\varepsilon}'_{uu}$ are large for uncertain boundary locations. Suppose $\mathbf{R}_u$ and $\mathbf{C}_{SS}$ are small, the uncertainty will be dominated by $\boldsymbol{\varepsilon}'_{uu}$. A similar case is considered by *Cardiff and Kitanidis* [2009], who used a gradient-based method to optimize the unknown facies boundary.
3. **$\mathbf{C}_{SS}$-dominated scenario.** This scenario represents the cases where the mean values of hydraulic properties of large-scale layers are known (i.e., $\mathbf{R}_u=\mathbf{0}$) as well as their distributions (i.e., $\boldsymbol{\varepsilon}'_{uu}=\mathbf{0}$). While the overall spatial mean and covariance of y at fine grids is m and a nested covariance $\mathbf{C}_{YY}$, the mean of each layer will be U_i for layer i if the geologic information is included. At the same time, the residual covariance of y at fine grids becomes $\mathbf{C}_{SS}$, with smaller variance and correlation scales. Such prior information has been utilized in HT analysis by *Tso et al.* [2016].

In real-world cases, $\mathbf{C}_{SS}$, $\mathbf{R}_U$, and $\boldsymbol{\varepsilon}'_{uu}$ likely exist at the same time. The proposed method is therefore more general, since it accounts for the three uncertainty sources simultaneously.

In order to further demonstrate and fully assess the proposed method, we next apply it to synthetic experiments, where exact large and small-scale K distributions are known exactly. Afterward, the proposed method is applied to a field problem, where a large-scale low permeability fault zone exists at a fractured granite site.

3. Synthetic Experiments

In this section, we use two synthetic numerical examples (i.e., examples 1 and 2) to demonstrate how site-specific geologic information (large-scale geologic pattern) is implemented through an initial mean parameter field and covariance functions in SLE, and to investigate how they affect the results of HT analysis. In example 1, we examine geologic media with relatively continuous large-scale patterns without distinct features. Example 2 is the case where a discrete feature (e.g., karst caverns or a large fault zone) exists.

3.1. Example 1

This synthetic example uses steady state HT to estimate the K distribution in a 2-D vertical sectional rectangular confined aquifer (45 m in length and 18 m in height), with a two-scale heterogeneous K field (Figure 2a, the reference K field). To generate this field, first, we discretize the aquifer into 15 $\times$ 18 coarse grids (3 m $\times$ 1m) and they are assigned with K values of a random field generated with a mean of 0.58 m/d, a ln$K$ variance of 1, and a horizontal correlation scale of 40 m and a vertical scale of 4 m with an exponential covariance function (i.e., $\mathbf{C}_{UU}$). A clustering method (i.e., k-means algorithm) [*Elsheikh et al.*, 2013] is subsequently applied to classify the K field into three types (facies) with mean ln$K$ values $U_1=-3.0$ (facies 1), $U_2=-1.0$ (facies 2), and $U_3=1.0$ (facies 3), as illustrated in Figure 2b. The facies map of the three types is shown in Figure 2c. The distribution of the three facies will be used in the following HT as our reference large-scale geologic pattern (or facies). Afterward, the aquifer is divided into 60 $\times$ 36 fine grids of 0.75 m $\times$ 0.5 m in size. The K value of each fine grid is generated using zero mean, variance of ln$K=0.2$ with correlation scales 6 m and 0.6 m in horizontal and vertical directions, respectively. Again, an exponential covariance function is used and it is denoted as $\mathbf{C}_{SS}$. These fine-grid K values are then added to the K field of coarse grids (the large-scale trend) to construct the true K field (Figure 2a).

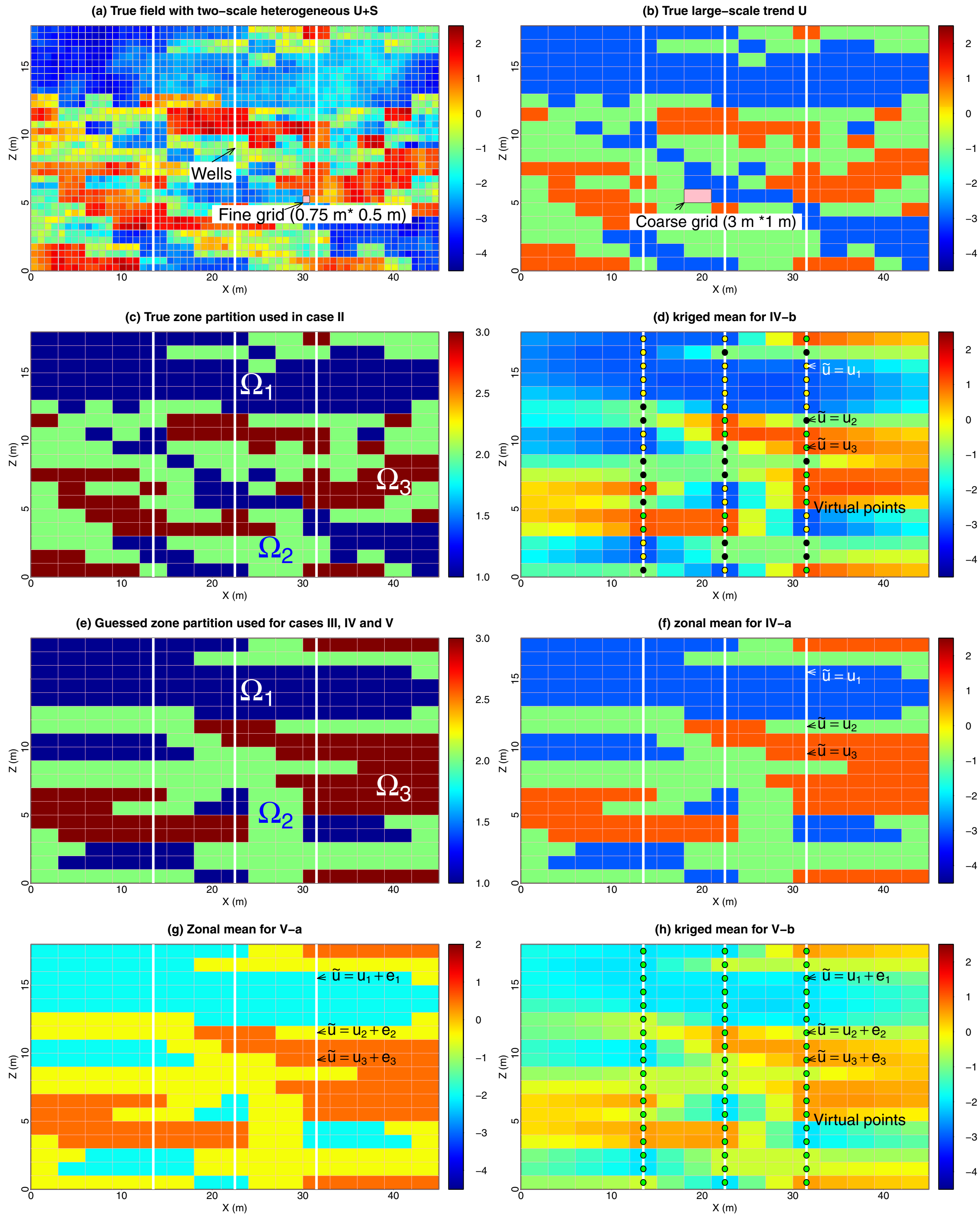

Figure 2. True field, large-scale trend, presumed facies geometry, and mean values used in example 1. The color indicates parameter value or facies number. (a) The true field *Y*. (b) Large-scale heterogeneity defined on the coarse grid. (c) The true zone geometry used in case II. (d) The kriged mean used for case IV-b. Kriging is based on the observations along the three boreholes without measurement error. (e) Guessed zone geometry (based on the kriged mean field) used in case III, IV, and V. (f) The zonal mean used for cases III and IV-a. (g) The zonal mean used for case V-a with an additional error. (h) The kriged mean used for case V-b, which is similar to Figure 2c with an additional error for kriging observations.

The HT test used in this example consists of three wells (Figure 2), and pumping is conducted at depths of 4, 8, 12, and 16 m of the wells. During each pumping test, steady state heads are collected at depths of 2, 4, 6, 8, 10, 12, 14, and 16 m in the three wells (excluding the pumping location). Thus, there are 12 tests and 23 observations in each test. We assume that the *K* values of the coarse grids and facies types along the three boreholes are known (with or without error).

Case I. This case represents the situation where HT interpretation uses a uniform mean (0.58 m/d) and the covariance $\mathbf{C}_{SS} + \mathbf{C}_{UU}$ as prior information (i.e., initial *K* field and its spatial structure). That is, we know only the generic information, but do not know any site-specific large-scale structure pattern, or facies distribution. The estimated *K* field is displayed in Figure 3a and its comparison with the reference *K* field is shown in the scatterplot in Figure 4a.

Case II. In this case, two scenarios (case II-a and case II-b) are considered. Both scenarios assume that the initial mean does not recognize the site-specific mean trend (distribution of low *K* and high *K*) and thus it uses the overall uniform mean ($m = \ln(0.58)$). However, the exact site-specific zone geometry of the reference field is implemented using covariance $\mathbf{R}_U$ (equation (21)), rather than the generic covariance $\mathbf{C}_{uu}$ (i.e., it is set zero). That is, based on the true facies distribution (Figure 2c), three different values, A_1, A_2, and A_3 are assigned to facies 1, 2, and 3 respectively, to form this $\mathbf{R}_U$. A_k is calculated based on the difference between assigned mean (m) and the zonal mean (U_k), i.e., $A_k = (U_k - m)^2$, where $k = 1$, 2, and 3. The actual values will affect the HT uncertainty estimation, but they play no role in the *K* estimate since it only depends on the spatial correlations, which are always ones and zeros in $\mathbf{R}_U$. In addition to this $\mathbf{R}_U$, Case II-a does not consider the covariance function of fine-grid properties, but case II-b does. In other words, $\varepsilon_{YY} = \mathbf{R}_U$ is used in case II-a, and $\varepsilon_{YY} = \mathbf{C}_{SS} + \mathbf{R}_U$ is used in case II-b. Figures 3b and 3c show the estimated *K* fields, and their comparisons with the reference *K* field are shown in the scatterplots in Figures 4b and 4c.

In case III, the site-specific large-scale pattern is estimated using coarse-grid *K* values along the three boreholes. Specifically, using these *K* values and $\mathbf{C}_{UU}$, kriging is employed to derive coarse-grid *K* values over the entire domain (Figure 2d). These *K* values are then converted to three patterns with different mean *K* values

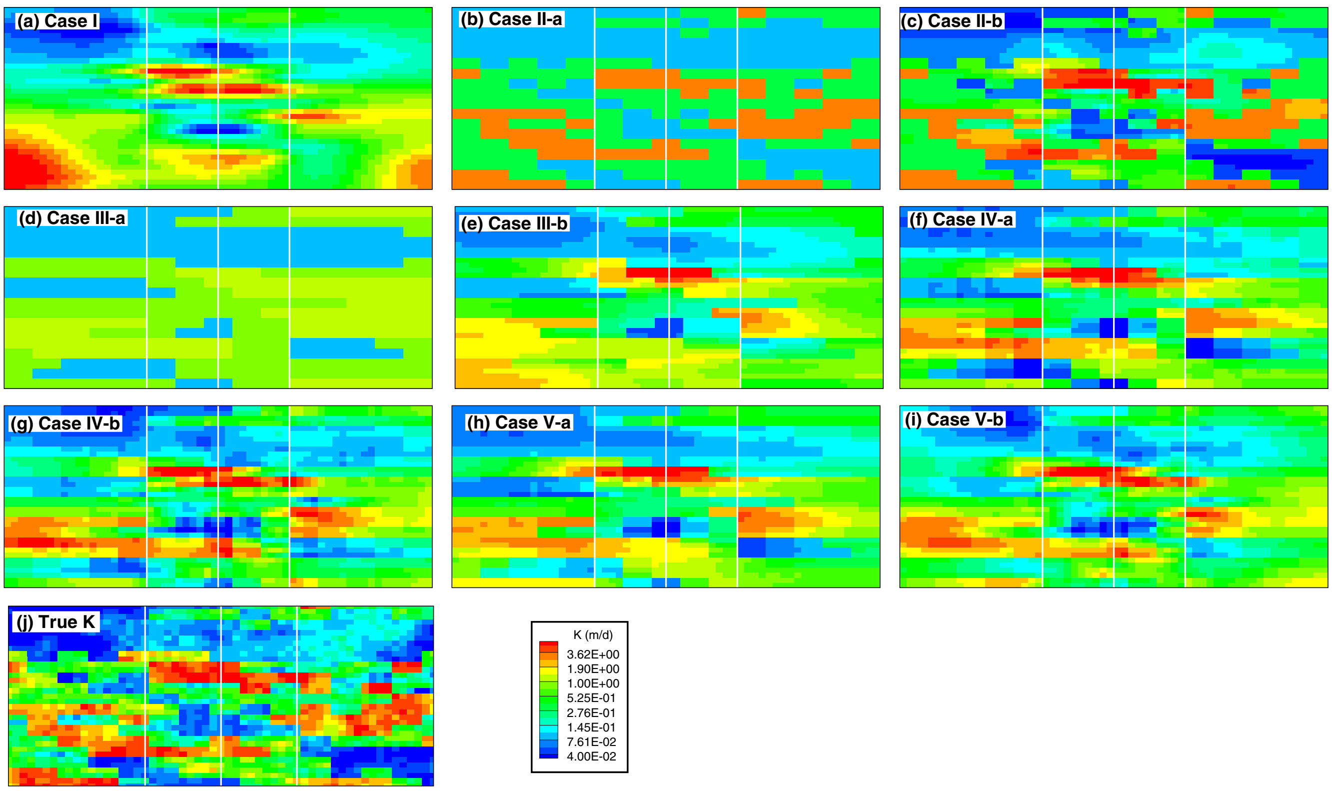

Figure 3. True *K* and estimated *K* fields from the nine scenarios of example 1.

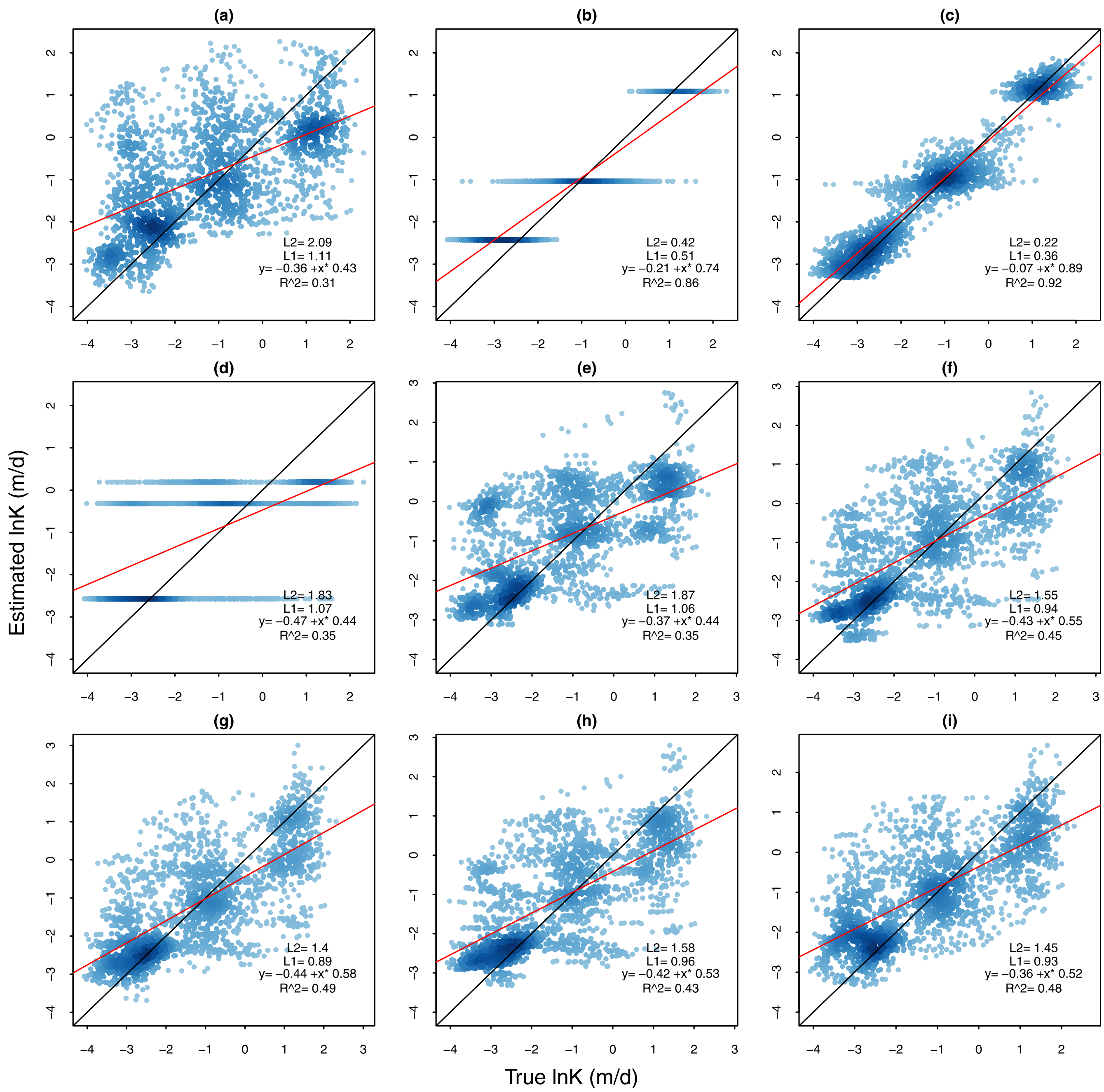

Figure 4. Scatterplots of true ln*K* versus estimated ln*K* in the nine scenarios of example 1. The darker color indicates a higher density of the scatters.

(Figure 2e), and three facies distributions (see Figure 2f) are defined according to their means. The generated distribution of the three facies is treated as our guessed large-scale pattern, and it is used to form a guessed $\mathbf{R}_U$ (denoted by $\hat{\mathbf{R}}_U$). This $\hat{\mathbf{R}}_U$ (with incorrect zone geometry) is different from the true $\mathbf{R}_U$ used in case II-a and case II-b (correct zone geometry).

In case III-a, we implement the large-scale spatial pattern using $\hat{\mathbf{R}}_U$, and let $\varepsilon_{YY}=\hat{\mathbf{R}}_U$ as the initial covariance to conduct HT analysis. On the other hand, in case III-b, we include the covariance of the fine grids and use $\varepsilon_{YY}=\mathbf{C}_{SS}+\hat{\mathbf{R}}_U$. Both case III-a and case III-b also use the uniform mean of the true field (0.58 m/d) as the initial mean *K*: it does not recognize any large-scale spatial pattern. The estimated *K* field is displayed in Figure 3d and its comparison with the reference *K* field is shown in the scatterplot (Figure 4d).

Next, we examine the case IV-a in which the initial mean K field is dictated by the incorrect facies distribution with correct mean K values for each facies type (thus, $\mathbf{R}_U = \mathbf{0}$) (Figure 2e), and the initial covariance only recognizes covariance of fine grid K. That is, $\mathbf{C}_{SS}$ is used as the initial covariance $\boldsymbol{\varepsilon}_{YY}$ for SLE. We also examine case IV-b in which the kriged mean K field (Figure 2d) is used as the initial mean K field, which is different from the true large-scale mean (Figure 2b). The covariance $\boldsymbol{\varepsilon}_{YY} = \boldsymbol{\varepsilon}'_{UU} + \mathbf{C}_{SS}$ is employed in case IV-b for SLE. Figures 3f and 3g show the estimated K fields and their comparisons with the reference K field are shown in the scatterplots in Figures 4f and 4g.

In the last part of this example, case V-a uses the incorrect facies pattern for the initial mean K field as that in case IV-a, and assigns incorrect K values ($U_1 = -2.0$, $U_2 = -0.5$, and $U_3 = 0.5$) for each facies (see Figure 2g). On the other hand, case V-b employs kriging to derive an initial mean K field similar to case IV-b, but it considers errors of the observations. This initial mean K field is illustrated in Figure 2h. Accordingly, case V-a uses $\boldsymbol{\varepsilon}_{YY} = \mathbf{C}_{SS} + \hat{\mathbf{R}}_U$ as the initial covariance, and case IV-b uses $\boldsymbol{\varepsilon}_{YY} = \boldsymbol{\varepsilon}'_{UU} + \mathbf{C}_{SS} + \hat{\mathbf{R}}_U$. Figures 3h and 3i show the estimated K fields and their comparisons with the reference K field are shown in the scatterplots in Figures 4h and 4i.

Examination of Figures 3 (estimated K field distributions) and 4 (scatterplots and associated performance statistics) shows that case II-b yields the best estimated K field. That is, using the covariance function (i.e., $\boldsymbol{\varepsilon}_{YY} = \mathbf{C}_{SS} + \mathbf{R}_U$) that reflects the true site-specific spatial pattern and recognizes spatial variability at the fine-grid, HT interpretation with SLE can lead to best estimates, in spite of the fact that the initial mean K field ignores the spatial pattern.

The runner-up is case II-a, which uses similar initial mean and covariance information as that in case II-b, but ignores the fine-grid covariance $\mathbf{C}_{SS}$. The third place belongs to case IV-b according to all performance statistics. This is the case which uses the kriged mean K field based on accurate coarse-grid K values of the observations along the three boreholes as the initial mean K field. Accordingly, the covariance becomes $\boldsymbol{\varepsilon}_{YY} = \boldsymbol{\varepsilon}'_{UU} + \mathbf{C}_{SS}$. The performances of the rest cases are in the following orders: case V-b, case IV-a, case V-a, case III-b, and lastly case I. Case IV-b and case V-b perform almost equally well and they do better than case III-b and case I.

These results show an interesting outcome. That is, using an initial covariance that reflects the correct site-specific large-scale pattern and fine-grid variability as prior information, SLE (or any Bayesian inverse model) can yield the best estimate in spite of the initial mean K field in the HT analysis. On the other hand, when these pieces of site-specific large-scale information are qualitative or imprecise, the results suggest that one uses directly or indirectly obtained K values and the kriging tool to derive the most likely large-scale K distribution and the associated residual covariance. Then, one should use this kriged K field and its residual covariance plus the fine-grid covariance in HT analysis such that it can lead to satisfactory results.

Overall, the site-specific covariance for the correct large-scale structure pattern is an additional constraint, which is more important than the initial mean K field. The proposed generalized geostatistical approach is versatile and valid, and it is a useful method for implementing the site-specific large-scale pattern when it is available.

3.2. Example 2

This example is to further strengthen the findings in example 1. Here, we consider a two-dimensional, rectangular synthetic confined aquifer (1000 m $\times$ 880 m), which is discretized into uniform elements with dimensions of 10 m in length and 8.8 m in width. The HT test consists of nine pumping tests at nine wells (white circles in Figure 5), and during each test, steady state heads are collected at the other eight wells. The reference synthetic K field is presented in Figure 5j, which consists of a distinct Z-shaped zone with a uniform low K value ($K_2 = 0.01$ m/d, denoting mean K of large-scale structure #2), surrounded by a high K zone ($K_1 = 1$ m/d, denoting mean K of large-scale structure #1). The high K zone is heterogeneous, described by an exponential covariance with isotropic correlation scale of 80 m and variance of $Y = 1$. The overall mean $m = 0.95$ m/d, large-scale variance $\sigma_U^2 = 0.9$, and correlation scales $L_U = 300$ m are used in the calculation. It should be noted that since the ergodicity condition is not met in this example, the large-scale covariance used here is not representative (see explanation in section 2.2.2).

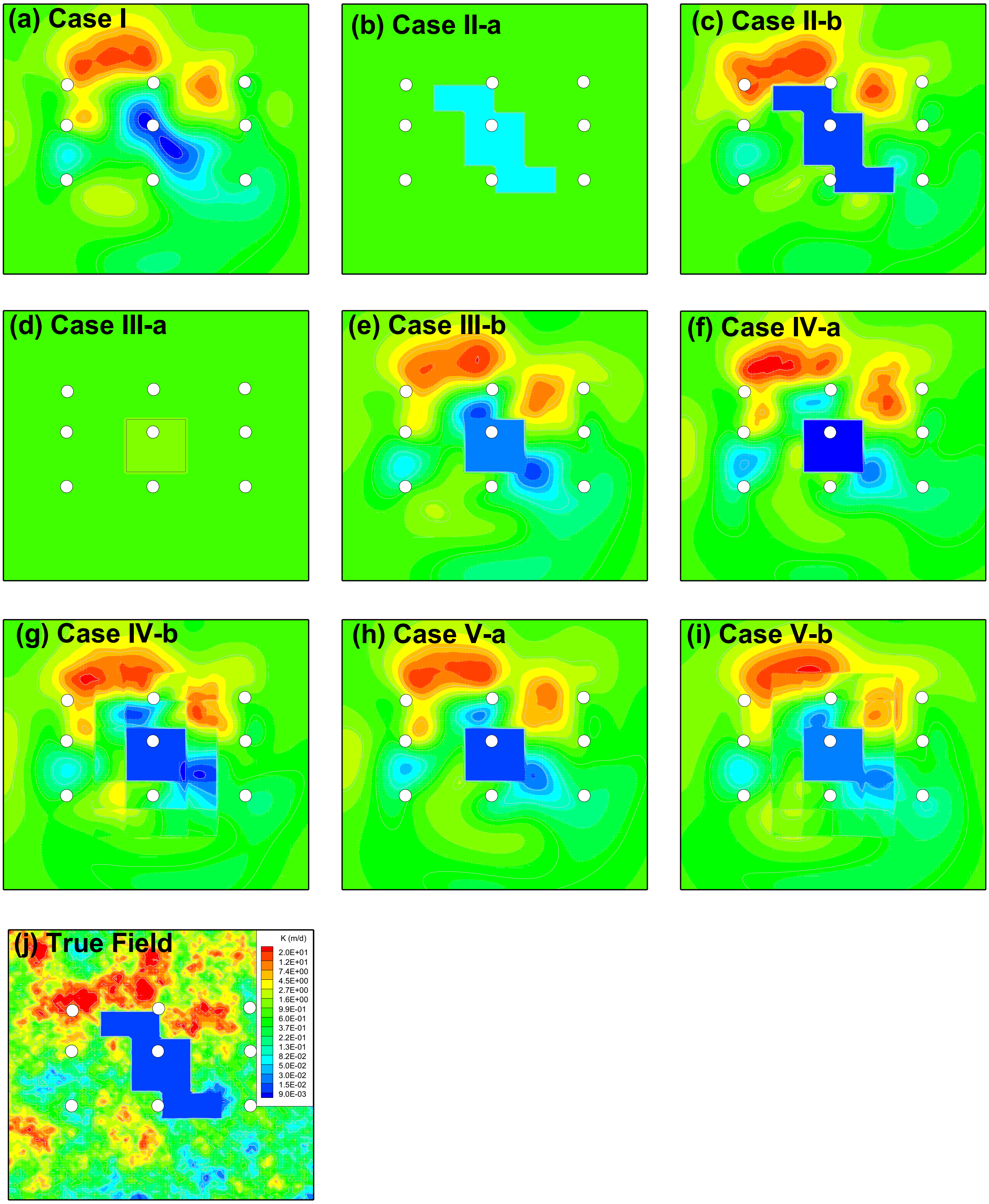

Figure 5. True K and estimated K field from the nine scenarios of example 2.

Table 1. Nine Scenarios Considering Different Inputs of Prior Information (Mean and Covariance)

Case	Scenario	Figure	Zone Geometry	Mean	Covariance
I[a]		Figures 3a, 4a, 5a, 6a	NA	Uniform	$\mathbf{C}_{UU}+\mathbf{C}_{SS}$
II[b]	a	Figures 3b, 4b, 5b, 6b	True	Uniform	$\mathbf{R}_U$
	b	Figures 3c, 4c, 5c, 6c	True	Uniform	$\mathbf{C}_{SS}+\mathbf{R}_U$
III[c]	a	Figures 3d, 4d, 5d, 6d	Guessed	Uniform	$\hat{\mathbf{R}}_U$
	b	Figures 3e, 4e, 5e, 6e	Guessed	Uniform	$\mathbf{C}_{SS}+\hat{\mathbf{R}}_U$
IV[d]	a	Figures 3f, 4f, 5f, 6f	Guessed	Zonal mean	$\mathbf{C}_{SS}$
	b	Figures 3g, 4g, 5g, 6g	Guessed	Kriged mean	$\boldsymbol{\varepsilon}'_{UU}+\mathbf{C}_{SS}$
V[e]	a	Figures 3h, 4h, 5h, 6h	Guessed	Zonal mean with error	$\mathbf{C}_{SS}+\hat{\mathbf{R}}_U$
	b	Figures 3i, 4i, 5i, 6i	Guessed	Kriged mean with error	$\boldsymbol{\varepsilon}'_{UU}+\mathbf{C}_{SS}+\hat{\mathbf{R}}_U$

[a]Using generic mean and covariance without site-specific information.
[b]Using true zone partition information and uniform mean.
[c]Using guessed zone partition (which is not the same as the true one) without consideration of possible boundary error and uniform mean.
[d]Using guessed zone partition. When possible boundary error is not considered, the covariance set $\boldsymbol{\varepsilon}'_{uu}=\mathbf{0}$ and zonal mean values are used. Otherwise, $\boldsymbol{\varepsilon}'_{uu}$ is included and kriged mean values are used.
[e]This is similar to case IV except that the mean value information contains errors.

Similarly, nine scenarios are considered to investigate different prior models (in terms of mean and covariance) on HT estimates (see Table 1). In case I, SLE uses the uniform mean K value and the two-scale covariance function, without including site-specific information, to estimate the detailed K distribution. Next, using the uniform mean K value, we investigate the effect of $\mathbf{R}_U$ in case II-a. Afterward, we use the uniform mean K value and both the $\mathbf{R}_U$ and $\mathbf{C}_{SS}$ in case II-b. With the same goals as those in scenarios case II-a and case II-b, case III-a and case III-b use incorrect site-specific covariance $\hat{\mathbf{R}}_U$ instead of $\mathbf{R}_U$. Case IV-a and case IV-b investigate the effect of incorrectly distributed true large-scale mean values. In contrast, scenarios case V-a and case V-b consider the error of the mean in the low K zone (low K zone value = 0.1 m/d compared to the true value 0.01 m/d). The virtual observations, large-scale grid (100 m × 88 m), kriged mean, and variance are provided in the supporting information Figure S1. The estimated K fields from different cases are plotted in Figure 5 and the scatterplots of simulated K versus true K of different scenarios are presented in Figure 6.

3.3. Summary of the Results of Examples 1 and 2

In summary, in both examples, a comparison of case I, case II-a, and case II-b, in which an initial uniform mean K field is used, demonstrates that inclusion of correct site-specific zone geometry information using covariance $\mathbf{R}_U+\mathbf{C}_{SS}$ has significant advantages compared to that using the generic large-scale covariance $\mathbf{C}_{UU}$. It maps both small-scale variability as well as large-scale distinct features rather clearly (e.g., case II-b in Example 2; Figure 5c). Results of case III-a and case III-b in both examples show that inclusion of $\mathbf{C}_{SS}$ still can improve the overall estimates even if the site-specific covariance $\hat{\mathbf{R}}_U$ is incorrect.

Under the condition of unknown zone geometry and correct virtual observations (case IV), using the kriged mean and covariance ($\boldsymbol{\varepsilon}'_{UU}$) yields a better result than the unconditional case (case I). Again, comparing case IV-a with case IV-b suggests that $\mathbf{C}_{SS}$ is helpful to identify small-scale features. In case V, the site-specific spatial patterns are the same as those in case IV, but their K values are incorrect due to measurement errors or indirect information about the hydraulic parameter. As expected, the additional error of the virtual observations in case V worsens the results compared to case IV, but it is still better than that of case I (without site-specific information). This indicates that inclusion of error-corrupted site-specific information (zone geometry and mean trend) can still improve HT results as long as its different uncertain sources are considered appropriately. One negative example is shown in cases II-a and III-a (Figures. 6b and 6e), where the results are much worse than unconditional case I (Figure. 6a) due to inappropriate ignorance of fine-grid variability ($\mathbf{C}_{SS}$). As indicated in Table 2, these scenarios that ignoring $\mathbf{C}_{SS}$ (II-a and III-a) often leads to bad quality of the head calibration. It is worthwhile to notice that including $\boldsymbol{\varepsilon}'_{UU}$ (the kriging variance of the coarse-grid parameter, representing uncertainty in the boundary of different facies or the distinct features) improves the K estimate in cases IV and V in example 1 (e.g., Figures. 4h and 4i). However, adding this term $\boldsymbol{\varepsilon}'_{UU}$ does not necessarily improve the estimate in example 2 (see Figures. 6h and 6i). This result may be attributed to the fact that $\boldsymbol{\varepsilon}'_{UU}$, based on the generalized geostatistical concept, is not appropriate for addressing uncertainty associated with discrete objects. That is, ergodicity is not met in this case (see explanation in section 2.2.2).

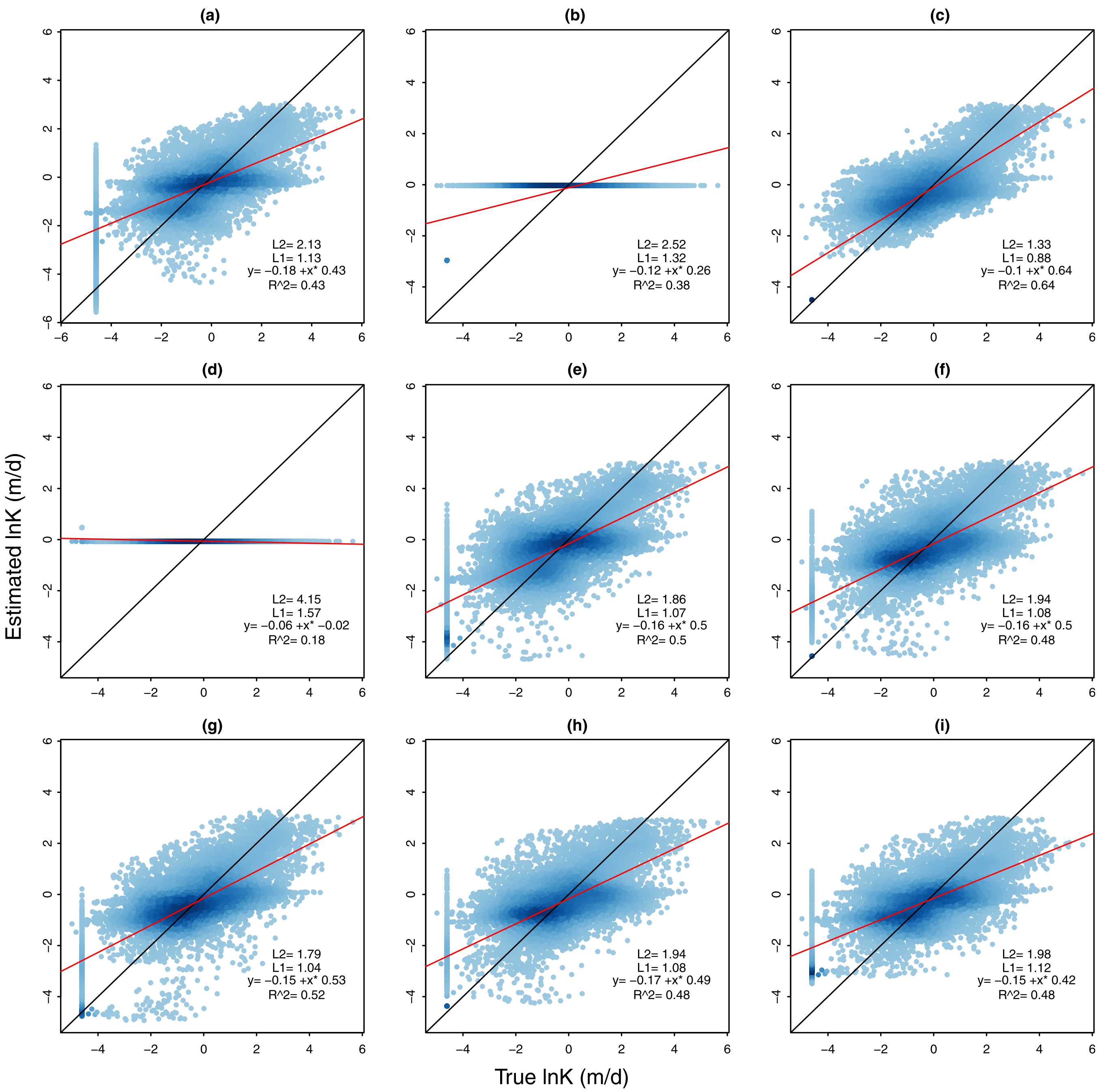

Figure 6. Scatterplots of true ln*K* versus estimated ln*K* in the nine scenarios of example 2. The darker color indicates a higher density of the scatters.

The two examples show that HT using the conditional mean and (residual) covariance derived from our generalized geostatistical framework consistently yields better result in different scenarios. It should be noted that the presented cases (cases II, III, IV, and IV) can be regarded as special scenarios under the general

Table 2. L2 Norms (m^2) of Simulated Head (Calibration Quality) for the Two Synthetic Cases

Case\Scenario	I	IIa	IIb	IIIa	IIIb	IVa	IVb	Va	Vb
Case 1	0.04	0.27	0.01	2.18	0.02	0.01	0.00	0.01	0.00
Case 2	0.00	0.18	0.00	0.23	0.00	0.00	0.00	0.00	0.00

framework of our proposed approach. For instance, if site-specific large-scale zone geometry from geologic mapping or geophysical surveys is available, zonal mean and residual covariance $\boldsymbol{\varepsilon}_{YY}=\mathbf{C}_{SS}+\mathbf{R}_U$ or $\boldsymbol{\varepsilon}_{YY}=\mathbf{C}_{SS}+\hat{\mathbf{R}}_U$ are used (cases II-b, III-b, and V-a). Implicitly, these cases place virtual observations on all the coarse grid and thus $\boldsymbol{\varepsilon}'_{UU}=\mathbf{0}$. Under the circumstances where site-specific zone distributions are not available, the proposed approach allows us to estimate the large-scale pattern using kriging with observations from well logs (case IV-b in example 1, see Figure 2d) or user-specified locations (case IV-b in example 2, see supporting information Figure S1f), and the derived conditional mean field and covariance $\boldsymbol{\varepsilon}_{YY}=\boldsymbol{\varepsilon}'_{UU}+\mathbf{C}_{SS}$ are used in HT. If the inaccuracy of mean values is considered besides the inaccurate site-specific spatial pattern, the proposed approach again uses kriging (case V-b in examples 1 and 2) and derives the conditional mean fields and the residual covariances ($\boldsymbol{\varepsilon}_{YY}=\boldsymbol{\varepsilon}'_{UU}+\mathbf{C}_{SS}+\hat{\mathbf{R}}_U$). The proportion of $\hat{\mathbf{R}}_U$ in $\boldsymbol{\varepsilon}_{YY}$ depends on the error involved in the virtual observations assigned by the users.

4. Field Application

In this section, we will apply the proposed approach to a field situation. Since the true field is not known, the estimated fields based on different prior information will be assessed via model validation. That is, the predictions of an independent test (which is not used in the inversion, or HT calibration) using different estimates are compared with field observed heads.

4.1. Description of the Field Site and Pumping Tests

During the past decade, Japan Atomic Energy Agency (JAEA) installed several vertical and inclined boreholes over an area of several square kilometers at depths of up to 1 km to characterize the hydrogeology near the Mizunami Underground Research Laboratory (MIU) site in central Japan. The site is situated in a fractured and faulted granite formation. A detailed description of the MIU site geology can be found in *Saegusa and Matsuoka* [2011]. According to *Saegusa and Matsuoka* [2011], geologic investigation suggested the existence of a prominent fault zone (flow barrier) that runs through the MIU site oriented North-North-West designated as fault IF_SB3_02 (Figure 7). The boundary between the Hongo Formation (AK/HG in) and the Toki Lignite-bearing Formation (TK) is also interpreted to be a flow barrier. Toki Granite, which underlies the Toki Lignite-bearing formation is highly fractured at depths between 300 and 500 m at MIU site. Beneath the highly fractured unit (upper highly fractured domain: UHFD) is a granitic body which is less fractured (lower sparsely fractured domain: LSFD), and is known to extend to great depths (Figure 7). However, the detailed hydrogeologic characteristics of these lineaments and faults are largely unknown.

At and around MIU site, nine vertical and slanted boreholes (MIZ-1, DH-2, DH-15, MSB-1, MSB-3, 10MI22, 05ME06, 07MI08, and 07MI09, see Figure 7 for their spatial locations) have been drilled, which have served as pumping or observation boreholes during the MIU site characterization effort. MIZ-1 penetrates to a depth of 1300 m, DH-15 is about 1000 m deep, DH-2 is about 500 m in depth, and the other boreholes are shallow boreholes with lengths between 100 and 200 m. DH-15 is situated approximately 500 m southeast from the MIU site. In contrast, the average distance between the off-site borehole DH-2 and all the other on-site boreholes are less than 200 m. Furthermore, when pumping at borehole 10MI22 (test 4, which will be introduced later), the newly added observation boreholes (05ME06, 07MI08, and 07MI09) are spaced less than 50 m around the main (MS) and ventilation shafts (VS) as shown in supporting information Table S1.

Three independent pumping tests (namely, tests 1, 2, and 4) were conducted at borehole MIZ1 and 10MI22, specifically for characterizing the hydrogeology of the site. In addition, during the course of excavation of the two vertical shafts (MS and VS), pumping at the shafts was administered to drain groundwater, which will be called test 3. Pumping locations of these four tests are shown in Figure 7. The observation boreholes were instrumented with multilevel monitoring systems. The pumping rates, durations, and observation boreholes and intervals associated with the four tests are listed in supporting information Table S2.

Test 1 was conducted from 14 to28 December 2004, at depth intervals of 191–226 m below the land surface along MIZ-1. The pumping location of test 2 was at a deeper depth interval (662–706 m) of the same borehole, and pumping test lasted from 13 to 28 January 2005. During these two tests, drawdowns were recorded in all observation intervals of boreholes DH-2, DH-15, MSB-1, and MSB-3.

Test 3 consisted of dewatering from the two shafts and responses were monitored at several depths along borehole MIZ-1 as well as at observation locations for tests 1 and 2. During the excavation of the shafts,

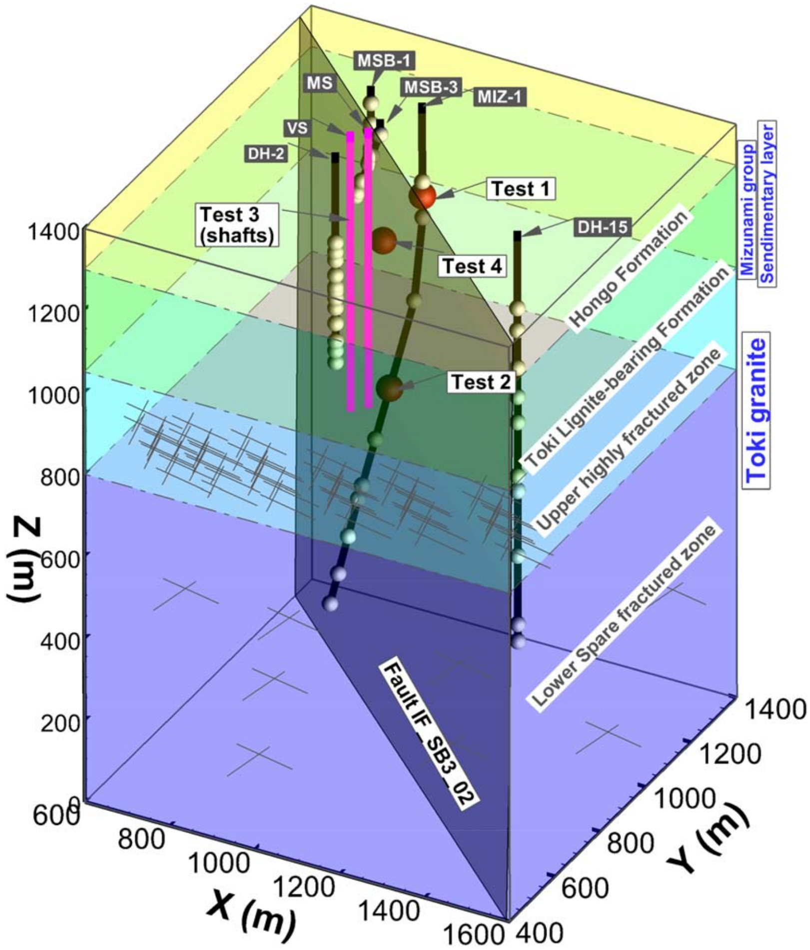

Figure 7. Geology, pumping locations, and observations intervals of MIU site. The locations of boreholes as well as the main shafts (MS) and ventilation shafts (VS) are shown. Pumping and observation intervals are indicated by spheres and the dash-dot lines approximately delineate the contact among various geologic units.

water is being drained out from the two shafts, thus the process is treated to be a pumping test with moving pumping interval.

During test 4, the pumping was conducted at the horizontal borehole 10MI22, which is 304 m below the surface and the pumping test started from 11 August 2010, lasting until 9 September 2010. Drawdowns at three additional boreholes (05ME06, 07MI08, and 07MI09) were monitored in addition to all previous observation locations. It was noticed that the continuous drainage from the shafts has created a large cone of depression, and has influenced the responses for test 4. Nevertheless, the rates of decrease in heads at all the observation intervals due to dewatering at the shafts were almost constant prior to test 4, since the drainage has been on-going for 4 years, reaching a quasi-steady state. Therefore, the estimated drawdown trends induced by the drainage of the shafts based on data prior to test 4 were removed from the drawdown-time data observed during test 4 (see details in supporting information). By doing this, we treated test 3 and test 4 as two independent tests.

4.2. Description of Hydraulic Tomography Analysis

The drawdown data from pumping tests 1 and 2 were analyzed by [*Illman et al*., 2009], and the data sets from tests 1, 2, 3, and 4 were later analyzed by *Zha et al.* [2015, 2016]. Both analyses did not consider known geologic structure information as prior information for their HT analysis. They started with a spatially uniform mean values of $K = 0.01$ m/d and $S_s = 2.3 \times 10^{-6}$ 1/m as well as an exponential covariance function with anisotropic correlation scales (i.e., 50, 50, and 25 m in x, y, and z directions for both K and S_s) while the variances were set to 2.0 for ln$K$ and 0.5 for lnS_s. Results of *Zha et al.* [2015, 2016] demonstrated that the inclusion of tests 3 and 4 increased the resolution of the estimated fracture zones and low-permeable fault zone, in particular the location, which seemed to agree with the general location observed through geologic investigation. However, the estimated low K zone is in a complex shape, indicating small-scale heterogeneity in the fault zone.

Based on the outcrop information of the fault and boreholes logs, the location and the orientation of the fault at the site are generally certain. The fault zone has been reported as a regional flow barrier. *Illman et al*. [2009] and *Zha et al*. [2015] have estimated that the fault zone has an average K of 0.003 m/d, approximately one order of magnitude lower than the K value of the matrix.

To implement this information (both structure and mean trend information) using our proposed approach for the following HT analysis, the fault and the matrix, as well as unknown fractures are considered as two distinct zones. In this case, we directly utilize the zone structure information without consideration of its uncertainty based on two reasons. First, the location and orientation of the fault zone are well constrained by the previous geologic survey. Second, as demonstrated in example 2, the benefit of including the boundary uncertainty is marginal if there are only several distinct zones, since small-scale heterogeneity ($\mathbf{C}_{ss}$) can compensate the incorrect mean due to wrong zone partition. Thus, here we assume that $\varepsilon'_{uu} = \mathbf{0}$. The kriged mean Y are assumed to be 0.003 m/d in the fault zone and 0.01 m/d in the matrix zone. Different means (0.001, 0.003, and 0.01 m/d) for the fault zone have been tried. The results indicate that the former two (0.001 and 0.003 m/d) produce similar predictions (see results below), but the latter one yields unimproved results (not shown here). To construct $\mathbf{R}_u$, $A_2 = 0.05$ is assigned for the fault zone and $A_1 = 0.5$ is set for the other zone, since the mean is relatively uncertain. These values are based on the residual variances of the HT inversion by *Zha et al*. [2015, 2016]. Moreover, $\mathbf{C}_{SS}$ is also constructed so that the heterogeneity inside the two zones are allowed. We assume a small-scale variance $\sigma_1^2 = 0.5$, isotropic correlation scale $L_1 = 20$ m in both the matrix and fractures, and $\sigma_2^2 = 0.05$ as well as an isotropic correlation scale $L_2 = 50$ m in the low K fault. For S_s, we start with a uniform mean (2.3×10^{-6} m^{-1}) and variability of lnS_s is described by a single covariance function with a variance of 0.5 and an isotropic correlation scale of 50 m. The mean and residual covariance used here is analogous to those used in case V-a of example 2.

4.3. Results of HT Analysis

To evaluate the effects of the proposed approach, test 4 is deliberately excluded from the inversion to serve as an independent test for the validation of HT estimates. That is, only data from tests 1 and 2 are used for HT inversion in the first two cases (Figures 8a and 8b), and the second two cases (Figures 9a and 9b) use data from tests 1, 2, and 3. Although both K and S_s are estimated in these cases, our discussion focuses on the resultant K estimates only since the estimated K and S_s are moderately negatively correlated [*Illman et al*., 2009]. The estimated S_s fields are provided in the supporting information section.

Figure 8a presents the estimated 3-D K distribution using data from tests 1 and 2 with a single covariance function and uniform means as prior information [see *Illman et al*., 2009; *Zha et al*., 2015, 2016]. It depicts the isosurfaces of the high and low K zones, and the contour K map of three horizontal slices (corresponding to the depths of tests 1, 4, and 2) are displayed in Figure 8c. As discussed in *Illman et al*. [2009] and *Zha et al*. [2015], the locations of tests 1 and 2 are on the same side of the low-permeability fault. The estimates reveal only one high K zone that connects DH-15 and MIZ-1 (borehole for tests 1 and 2). The low K zone spreads over a wide area including the upper parts of MSB-1, MSB-3, the shafts, and borehole DH-2.

Figure 8b illustrates the estimates using the same data set as that in Figure 8a, except that the geologic knowledge of the fault location is incorporated by our proposed approach as prior information during HT analysis. It shows that the high K zone, which connects DH-15 and MIZ-1, remains largely unchanged as in Figure 8a. Nevertheless, the shape of the low K zone is altered significantly as illustrated in Figures 8b and d. This zone stretches out horizontally and vertically to become a fault plane, which is quite different from that in Figure 8a. The majority of the presumed fault zone via prior information remains, since the groundwater response observations are sparse and only two tests are available; they are not able to override the prior information. As a consequence, the resultant low K zone is much sharper than that in Figure 8a.

AGU Water Resources Research 10.1002/2016WR019185

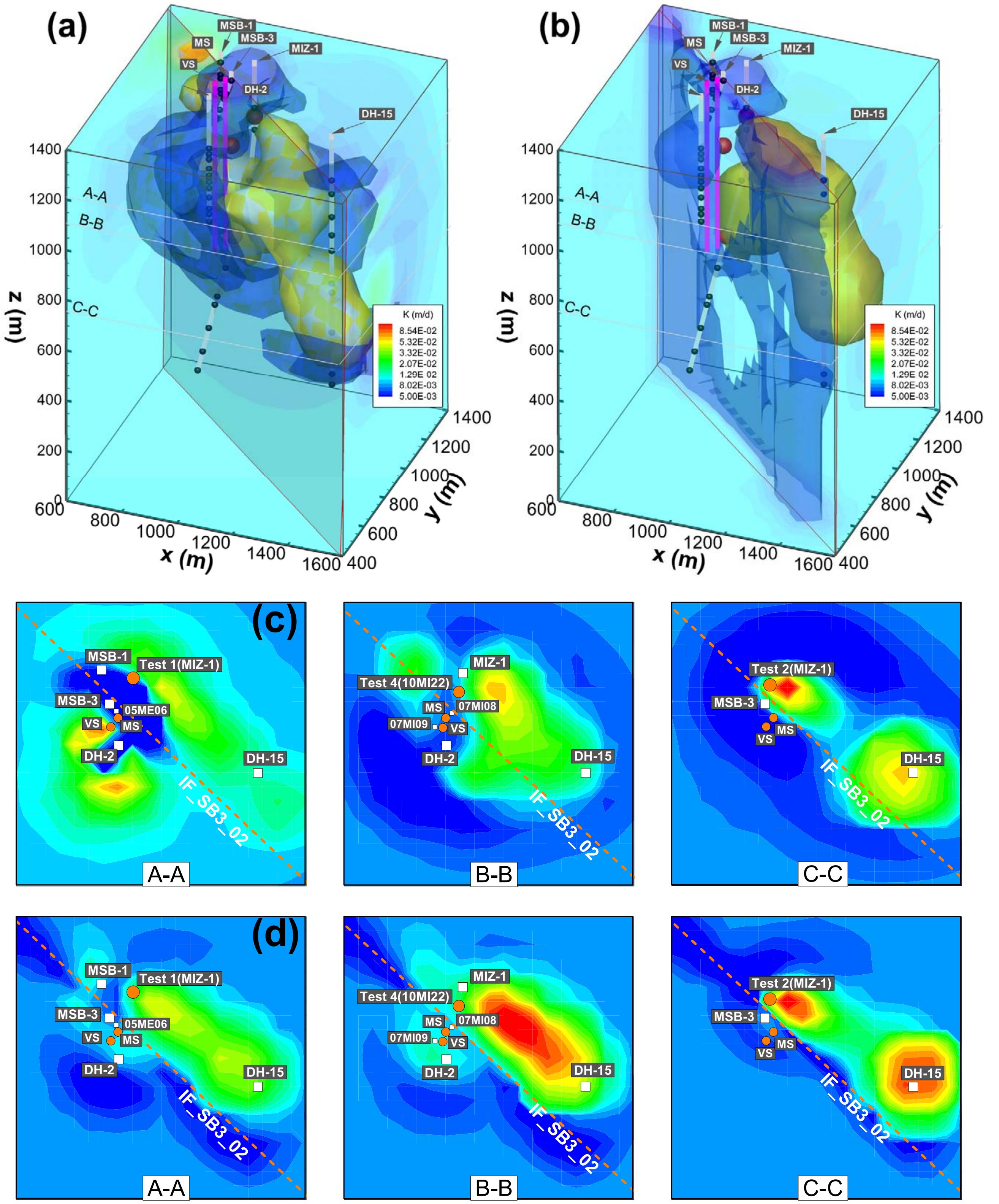

Figure 8. *K* estimates using data from tests 1 and 2 (a) without or (b) with prior knowledge of the low permeable fault. (c) And (d) are three slices from 3-D estimates in Figures 8a and 8b, respectively. Both estimates show that the pumping locations of tests 1-2 (upper and lower intervals of MIZ-1) are connected to borehole DH-15 but are isolated from boreholes DH-2. The location of test 4 (10MI22 in B-B, not used in the inversion) is connected to DH-15 in (b) and less connected in Figure 8a.

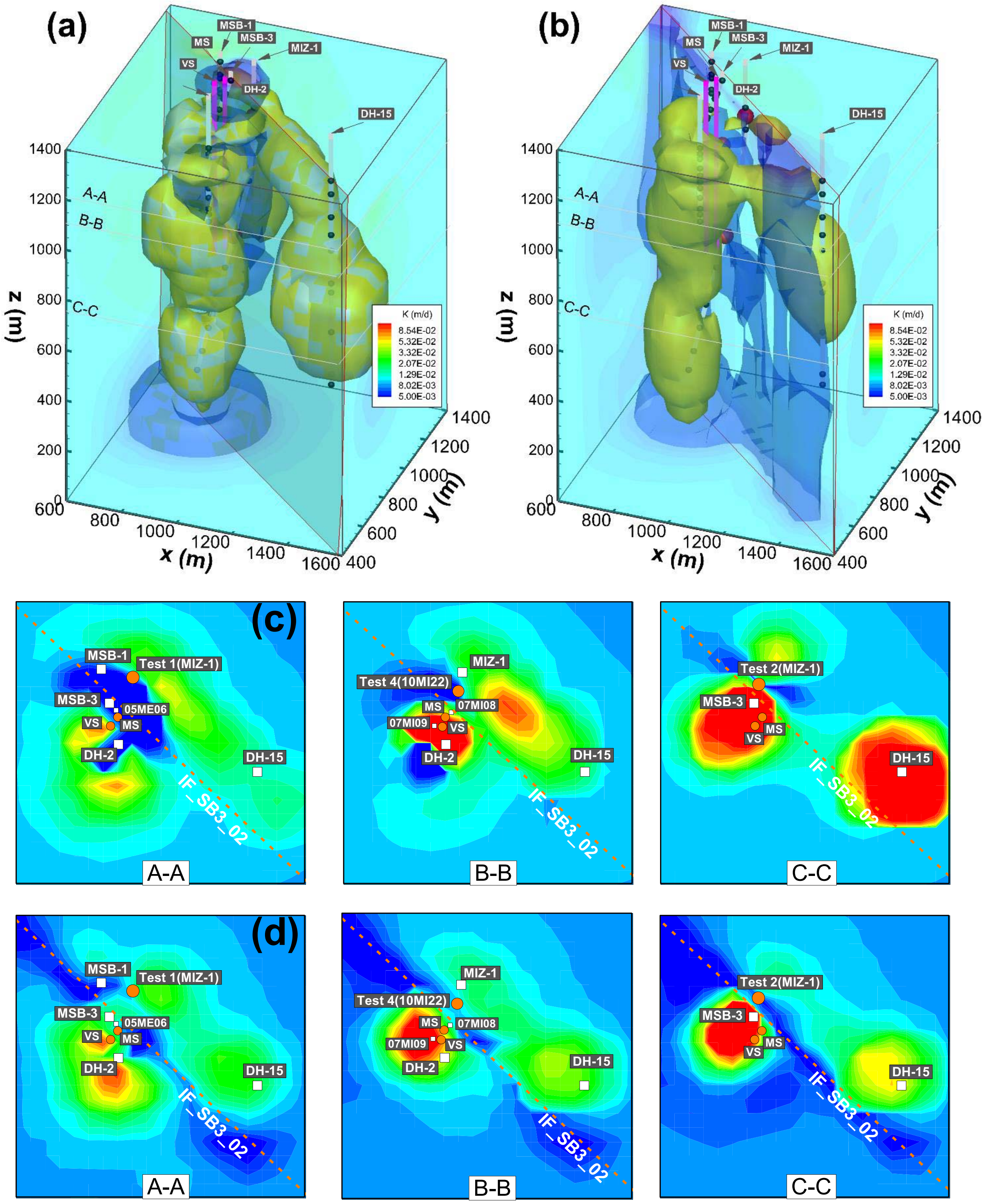

Figure 9. *K* estimates using data from tests 1, 2, and 3 (a) without or (b) with prior knowledge of the low permeable fault. (c) And (d) are three slices from 3-D estimates in Figures 9a and 9b, respectively. Both estimates show that the pumping locations of tests 1-2 (upper and lower intervals of MIZ-1) are connected to borehole DH-15 while boreholes DH-2 is connected to the two shafts (test 3). The location of test 4 (10MI22 in B-B, not used in the inversion) is not connected to DH-15 in Figure 9a but connected in Figure 9b.

HT inverse results based on data from tests 1, 2, and 3 with a uniform mean and a single covariance as prior information are illustrated in Figure 9a and the *K* contour maps of associated three horizontal slices are displayed in Figure 9c. One significant difference between Figures 8 and 9 is that the new result reveals a new high *K* zone that connects DH-2 and the two shafts. This finding is attributed to the fact that test 3 was conducted on the other side of the fault compared to tests 1 and 2 [*Zha et al*., 2015]. Due to this additional non-redundant data, the estimated low *K* zone becomes narrower between DH-2 and MIZ-1 at some depths. However, as illustrated in B-B horizontal slices in Figures 9c, the pumping location of test 4 (10MI22, which is not used in the inversion), now is located in the low *K* zone.

On the other hand, employing the geologic knowledge of the fault location as prior information, the inverse results (Figures 9b and 9d) reveal that 10MI22 is connected to DH-15. This refinement may not be significant in the kilometer-scale view of the *K* estimates, but it has a great impact on the validation results of test 4, since it changes the connectivity or lack of it between the new pumping location and other observation intervals.

The above results in Figures 8 and 9 indicate that both HT data and prior information impact the final estimates. This is consistent with the Bayesian nature of SLE algorithm used in the inversion. The estimates show that the prior mean values are retained at some locations (especially those far away from the well field) after inversion, because no observed head data are close by to improve the prior estimates. Furthermore, we find that the prior geologic information leads to different patterns of heterogeneity inside the well field (e.g., the shape of the low *K* zone, and the connectivity of DH-15 and 10MI22). This is likely caused by the fact that the cross-variances (used to update the parameter fields) are different and there are no observed data that are close to overrule them.

Figure 10 illustrates the reduction of parameter ln*K* uncertainty for the inversions using data from pumping tests 1, 2, and 3, with or without using the fault information as the prior information. The reduction of uncertainty at every location in the domain is defined as the relative change of the residual variance (diagonal term of the covariance matrix calculated by equation (4)) from the unconditional variance (user-specified value without using the fault information or ε_{YY} in equation (24) when using the fault information). The reduction in uncertainty of parameter ln*K* is significant in the region near the pumped well and the observation intervals. Figures 10b and 10d show larger uncertainty reductions along the fault position compared to that in Figures 10a and 10c. This is due to the fact that Figures 10b and 10d use distributed mean and non-stationary prior covariance to include fault information. Moreover, comparing Figure 10a with 10b, we see that the uncertainty is reduced in the deep region where test 2 is conducted when the fault information is included in the inversion. *Zha et al*. [2016] also reported that the inclusion of more tests decreases the uncertainty significantly.

4.4. Validation

One way to assess the improvements of the estimate due to the inclusion of the prior geostatistical information is to compare the estimate with the true *K* distribution at the site, but it is unknown and thus impossible. Alternatively, they can be assessed by validation. That is, we use the estimated parameter fields to predict the flow fields induced by test 4, which has not been used in the inversion. Then, comparisons of the predicted and observed drawdown at all observation ports during test 4 in a scatter plot with statistical norms calculated by equations (12) and (13) would be able to quantitatively show any improvements.

Figure 11a is the scatter plot of the observed and predicted drawdown based on the estimates from the inversion of tests 1-2 without prior information about the fault (i.e., estimates in Figure 8a). Generally speaking, the L1 (2.38 m) and L2 (9.53 m^2) norms indicate that the validation is poor. The drawdown behaviors at borehole DH-15 is quite accurately predicted (blue dots in the figure), while the predicted drawdowns at all the other observation intervals are much greater than the observed. Thus, the prediction is highly biased. As the information of fault is included as prior information, the L1 and L2 (Figure 11b) become smaller than those in Figure 11a indicating that the inclusion of fault information improves the estimates slightly.

The predicted drawdown using the estimate (Figure 9a) of the inversion for tests 1, 2, and 3, without using prior information about the fault is plotted against the observed in Figure 11c. Comparing this result to those in Figures 11a and 11b, we notice that the predicted drawdown has lower L1 and L2 norms (0.70 m and 0.92 m^2) and is relatively unbiased after the inclusion of the drawdown data from test 3. Nevertheless,

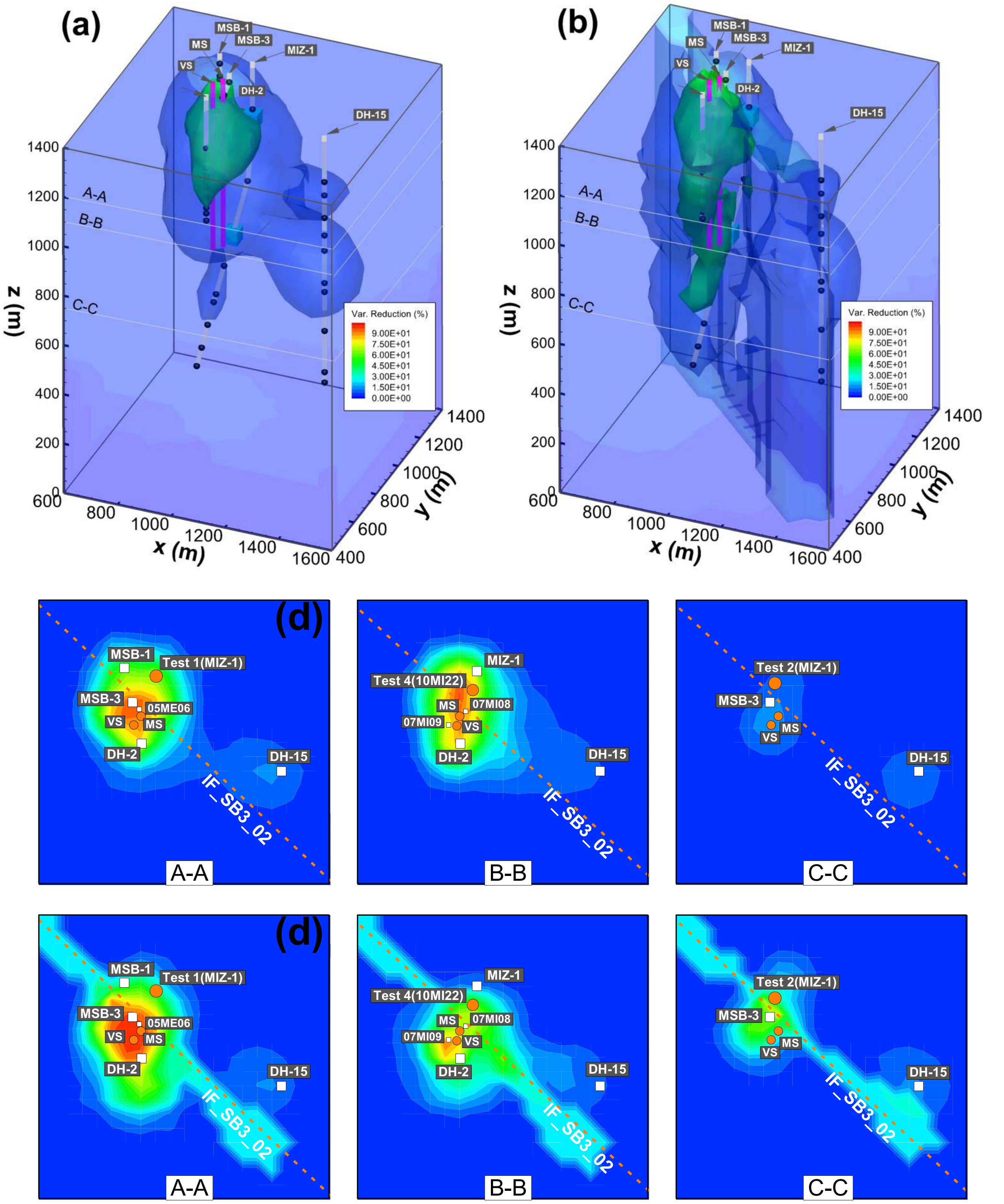

Figure 10. Variance reduction (%) of the estimated *K* using data from tests 1, 2, and 3 (a) without or (b) with prior knowledge of the low permeable fault. (c) And (d) are three slices from 3-D estimates in Figures 10a and 10b, respectively.

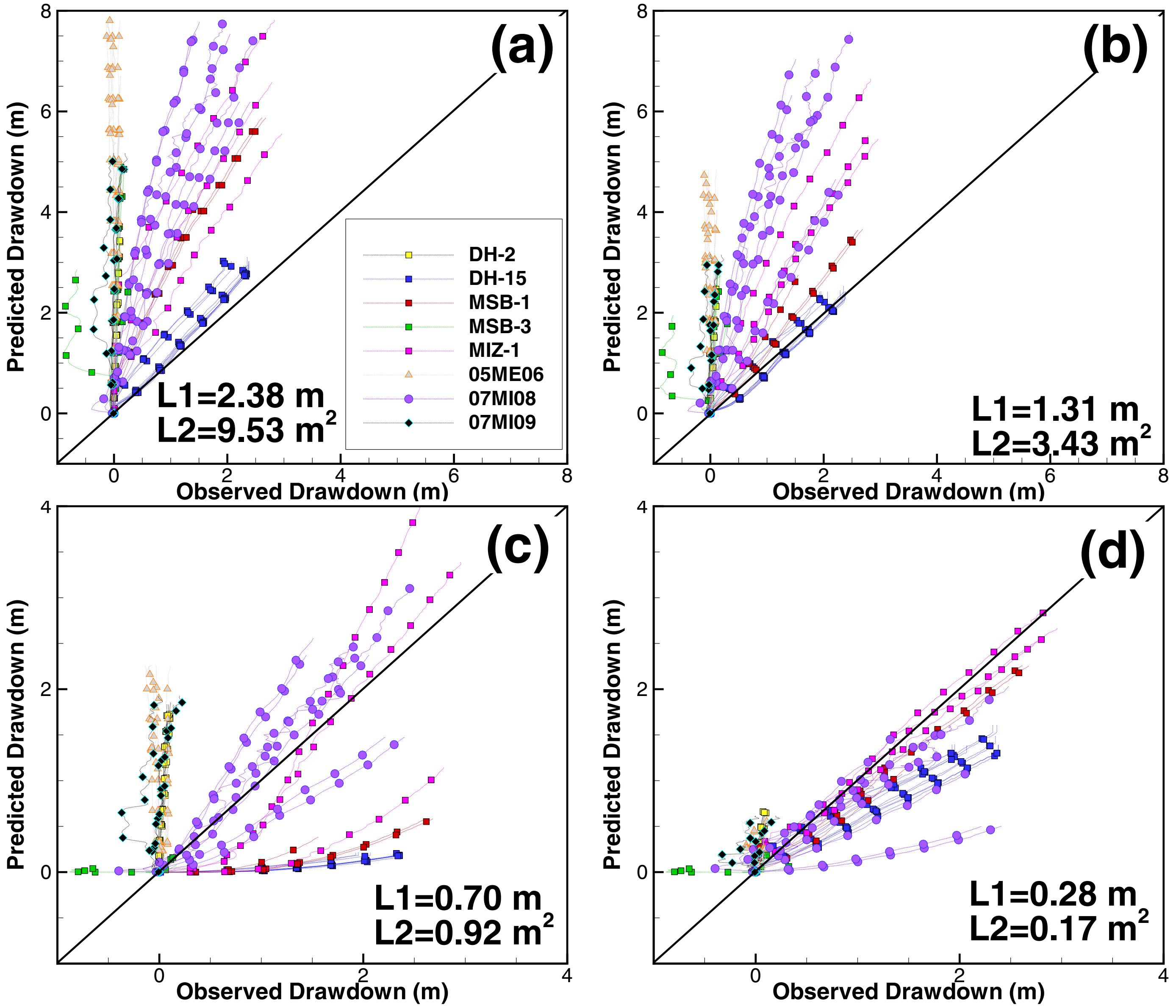

Figure 11. Validation results of the estimates using the HT data of tests 1 and 2 (a) without and (b) with prior geologic information (b); validation results of the estimates using 1, 2, and 3 (c) without and (d) with prior geologic information. The scales are different in each plot.

the scatter remains large. In particular, the agreement between the predicted and observed responses in borehole DH-15 (blue dots) worsens. This is likely a result of the misinterpretation of the connection between DH-15 and the pumping location for test 4, 10MI22 as illustrated in the estimates shown in Figures 9a and 9c.

Once the prior information of the fault, in addition to the drawdown data from test 3, is included to derive the HT estimate (Figure 9d), the estimate yields unbiased predictions with much smaller L1 and L2 norms in Figures 11d. In particular, the predicted responses in borehole DH-15 are much closer to those observed. This validation result is a clear indication of improvements on the estimated parameter fields due to incorporating prior geologic information in HT analysis.

A comparison of Figures 11a and 11c suggests that the inclusion of test 3 data has significant impacts on the estimate and in turns, the predicted drawdown. We notice that the pumping location of test 3 is located on the west side of the fault, while tests 1 and 2 are on the east side of the fault. That is, data from test 3 allow HT to identify the fracture connectivity on the west side of the fault, which was not revealed by data from tests 1 and 2 (see Figures 8 and 9). Further consideration of the prior geostatistical information during HT analysis then refines the estimate, which yields much better validation results.

5. Conclusions

The revisit of SLE and MAP algorithms demonstrates that both SLE and MAP are built upon the Bayesian framework, in which covariance is an additional soft constraint. Nevertheless, SLE continues to update estimates and the soft constraint (residual covariance) during iterations as opposed to MAP. The update of the residual covariance is essential: it reflects the fact that additional information about the properties is extracted from observed responses, which improve the estimate and corrects the soft constraint (residual covariance) during each iteration.

The concept of updating the residual covariance leads to the development of a general geostatistical approach to incorporate a hierarchical residual covariance, which can reflect the site-specific large-scale geologic patterns as well as small-scale variability. Specifically, heterogeneity at a field site is conceptualized as a random field with two-scale nested generic covariance function. To tailor such a generic description of the variability of a large-scale property to a site-specific one, kriging is applied. It uses observed attributes to derive the most likely spatial distribution of the property given virtual observations assigned based on site-specific geologic information. The resultant (residual) covariance reflects the estimated uncertainties from measurement error, indirect observation, and imprecise large-scale spatial pattern. This residual covariance subsequently serves as prior geologic information for any Bayesian inverse model. This approach is versatile; it can consider cases where site-specific large-scale geologic patterns are known or unknown; it also considers fine-scale variability within facies or stratifications.

We use numerical experiments to demonstrate the usefulness of this framework in two cases with two different types of heterogeneity. We demonstrate that with the help of proposed framework, site-specific large-scale information can be assimilated and improve HT results. The results corroborate our hypothesis that prior information (in terms of mean distribution and site-specific covariance function) can improve the estimates of HT surveys using the SLE algorithm when the spatial observations of aquifer responses are sparse.

Last, the proposed approach with SLE is applied to the analysis of HT surveys conducted at a fractured and faulted granite field site. The robustness of the estimates is assessed by their ability to predict observed drawdowns induced by an independent pumping test, not utilized for the estimation. Comparison of observed and predicted drawdowns confirms that the proposed approach indeed improves the estimates and in turn, the prediction. Nonetheless, the improvement is most notable when an appropriate layout of the pumping well in HT survey is used. This result supports the conclusions of studies by *Zha et al.* [2014, 2015, 2016] that pumping tests must be conducted on both sides of a low-permeable fault zone at this site. The result also substantiates findings by *Yeh and Liu* [2000] and *Liu et al.* [2002] about the importance of collecting sufficient data during HT surveys.

Acknowledgments
The work was supported by the U.S. Environmental Security Technology Certification Program (ESTCP) grant ER-201212. Additional support partially comes from the NSF EAR grant 1014594. The first author acknowledges the supports from National Natural Science Foundation of China (51609173) and the open-end fund provided by Key Laboratory for Groundwater and Ecology in Arid and Semi-Arid Areas, CGS (KLGEAS201601). J.-C. Wen would like to acknowledge the research support from MOST 103-2221-E-224-054, MOST 104-2221-E-224-039, and MOST 105-2625-M-224-002 by the Minister of Science and Technology, Taiwan. T.-C. J. Yeh also acknowledges the Outstanding Overseas Professorship award through Jilin University by the Department of Education, China, as well as the Global Expert award through Tianjin Normal University from the Thousand Talents Plan of Tianjin City. Walter A. Illman acknowledges the support from the Discovery Grant of the Natural Sciences & Engineering Research Council of Canada. All data from this work are available upon request through the corresponding author. The constructive suggestions from the associate editor and three reviewers are appreciated.

References

Brauchler, R., J.-T. T. Cheng, P. Dietrich, M. Everett, B. Johnson, R. Liedl, and M. Sauter (2007), An inversion strategy for hydraulic tomography: Coupling travel time and amplitude inversion, *J. Hydrol.*, *345*(3–4), 184–198, doi:10.1016/j.jhydrol.2007.08.011.

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.

Cardiff, M., W. Barrash, and P. K. Kitanidis (2012), A field proof-of-concept of aquifer imaging using 3-D transient hydraulic tomography with modular, temporarily-emplaced equipment, *Water Resour. Res.*, *48*, W05531, doi:10.1029/2011WR011704.

Carrera, J., and L. Glorioso (1991), On geostatistical formulations of the groundwater flow inverse problem, *Adv. Water Resour.*, *14*(5), 273–283, doi:10.1016/0309-1708(91)90039-Q.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L. J. Slooten (2005), Inverse problem in hydrogeology, *Hydrogeol. J.*, *13*(1), 206–222, doi:10.1007/s10040-004-0404-7.

Chen, Y., and D. S. Oliver (2013), Levenberg–Marquardt forms of the iterative ensemble smoother for efficient history matching and uncertainty quantification, *Comput. Geosci.*, *17*(4), 689–703, doi:10.1007/s10596-013-9351-5.

Dietrich, C. R., and G. N. Newsam (1989), A stability analysis of the geostatistical approach to aquifer transmissivity identification, *Stoch. Hydrol. Hydraul.*, *3*(4), 293–316, doi:10.1007/BF01543462.

Elsheikh, A. H., M. F. Wheeler, and I. Hoteit (2013), Clustered iterative stochastic ensemble method for multi-modal calibration of subsurface flow models, *J. Hydrol.*, *491*(1), 40–55, doi:10.1016/j.jhydrol.2013.03.037.

Eppstein, M. J., and D. E. Dougherty (1996), Simultaneous estimation of transmissivity values and zonation, *Water Resour. Res.*, *32*(11), 3321–3336, doi:10.1029/96WR02283.

Fienen, M., R. Hunt, D. Krabbenhoft, and T. Clemo (2009), Obtaining parsimonious hydraulic conductivity fields using head and transport observations: A Bayesian geostatistical parameter estimation approach, *Water Resour. Res.*, *45*, W08405, doi:10.1029/2008WR007431.

Fienen, M. N., T. Clemo, and P. K. Kitanidis (2008), An interactive Bayesian geostatistical inverse protocol for hydraulic tomography, *Water Resour. Res.*, *44*, W00B01, doi:10.1029/2007WR006730.

Hendricks Franssen, H. J., A. Alcolea, M. Riva, M. Bakr, N. van der Wiel, F. Stauffer, and A. Guadagnini (2009), A comparison of seven methods for the inverse 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.

Grana, D., M. Pirrone, and T. Mukerji (2012), Quantitative log interpretation and uncertainty propagation of petrophysical properties and facies classification from rock-physics modeling and formation evaluation analysis, *Geophysics*, *77*(3), WA45, doi:10.1190/geo2011-0272.1.

Huang, S.-Y., J.-C. Wen, T.-C. J. Yeh, W. Lu, H.-L. Juan, C.-M. Tseng, J.-H. Lee, and K.-C. Chang (2011), Robustness of joint interpretation of sequential pumping tests: Numerical and field experiments, *Water Resour. Res.*, *47*, W10530, doi:10.1029/2011WR010698.

Hughson, D., and T. Yeh (2000), An inverse model for three-dimensional flow in variably saturated porous media, *Water Resour. Res.*, *36*(4), 829–839, doi:10.1029/2000WR900001.

Illman, W. A. (2014), Hydraulic tomography offers improved imaging of heterogeneity in fractured rocks, *Groundwater*, *52*(5), 659–684, doi:10.1111/gwat.12119.

Illman, W. A., X. Liu, S. Takeuchi, T.-C. J. Yeh, K. Ando, and H. Saegusa (2009), Hydraulic tomography in fractured granite: Mizunami underground research site, Japan, *Water Resour. Res.*, *45*, W01406, doi:10.1029/2007WR006715.

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

Konikow, L. F., and J. D. Bredehoeft (1992), Ground-water models cannot be validated, *Adv. Water Resour.*, *15*(1), 75–83, doi:10.1016/0309-1708(92)90033-X.

Liu, S., T.-C. J. Yeh, and R. Gardiner (2002), Effectiveness of hydraulic tomography: Sandbox experiments, *Water Resour. Res.*, *38*(4), 1034, doi:10.1029/2001WR000338.

Liu, X., Q. Zhou, P. K. Kitanidis, and J. T. Birkholzer (2014), Fast iterative implementation of large-scale nonlinear geostatistical inverse modeling, *Water Resour. Res.*, *50*, 198–207, doi:10.1002/2012WR013241.

McLaughlin, D., and L. R. Townley (1996), A reassessment of the groundwater inverse problem, *Water Resour. Res.*, *32*(5), 1131–1161, doi:10.1029/96WR00160.

Meier, P., A. Medina, and J. Carrera (2001), Geostatistical inversion of cross-hole pumping tests for identifying preferential flow channels within a shear zone, *Groundwater*, *39*(1), 10–17, doi:10.1111/j.1745-6584.2001.tb00346.x.

Ni, C., T. Yeh, and J. Chen (2009), Cost-effective hydraulic tomography surveys for predicting flow and transport in heterogeneous aquifers, *Environ. Sci. Technol.*, *43*(10), 3720–3727.

Nowak, W., and O. A. Cirpka (2004), A modified Levenberg-Marquardt algorithm for quasi-linear geostatistical inversing, *Adv. Water Resour.*, *27*(7), 737–750, doi:10.1016/j.advwatres.2004.03.004.

Poeter, E. P., and S. A. McKenna (1995), Reducing uncertainty associated with ground-water flow and transport predictions, *Groundwater*, *33*(6), 899–904, doi:10.1111/j.1745-6584.1995.tb00034.x.

Pujol, J. (2007), The solution of nonlinear inverse problems and the Levenberg-Marquardt method, *Geophysics*, *72*(4), W1, doi:10.1190/1.2732552.

Ronayne, M. J., S. M. Gorelick, and J. Caers (2008), Identifying discrete geologic structures that produce anomalous hydraulic response: An inverse modeling approach, *Water Resour. Res.*, *44*, W08426, doi:10.1029/2007WR006635.

Saegusa, H., and T. Matsuoka (2011), Final report on the surface-based investigation phase (phase 1) at the Mizunami Underground Research Laboratory project, JAEA-Research 2010-067, Ibaraki, Japan.

Soueid Ahmed, A., J. Zhou, A. Jardani, A. Revil, and J. P. Dupont (2015), Image-guided inversion in steady-state hydraulic tomography, *Adv. Water Resour.*, *82*, 83–97, doi:10.1016/j.advwatres.2015.04.001.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, Pa.

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.

Tso, C.-H. M., Y. Zha, T.-C. J. Yeh, and J.-C. Wen (2016), The relative importance of head, flux, and prior information in hydraulic tomography analysis, *Water Resour. Res.*, *52*, 3–20, doi:10.1002/2015WR017191.

Xiang, J., T.-C. J. Yeh, C.-H. Lee, K.-C. Hsu, and J.-C. Wen (2009), A simultaneous successive linear estimator and a guide for hydraulic tomography analysis, *Water Resour. Res.*, *45*, W02432, doi:10.1029/2008WR007180.

Yeh, T. -C., D. Mao, Y. Zha, J. Wen, L. Wan, K. Hsu, and C. Lee (2015a), Uniqueness, scale, and resolution issues in groundwater model parameter identification, *Water Sci. Eng.*, *8*(3), 175–194, doi:10.1016/j.wse.2015.08.002.

Yeh, T.-C., R. Khaleel, and K. C. Carroll (2015b), *Flow Through Heterogeneous Geologic Media*, 343 pp., Cambridge Univ. Press, Cambridge, U. K.

Yeh, T.-C. J., and S. Liu (2000), Hydraulic tomography: Development of a new aquifer test method, *Water Resour. Res.*, *36*(8), 2095–2105, doi:10.1029/2000WR900114.

Yeh, T.-C. J., M. Jin, and S. Hanna (1996), An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields, *Water Resour. Res.*, *32*(1), 85–92.

Zha, Y., T.-C. J. Yeh, D. Mao, J. Yang, and W. Lu (2014), Usefulness of flux measurements during hydraulic tomographic survey for mapping hydraulic conductivity distribution in a fractured medium, *Adv. Water Resour.*, *71*, 162–176, doi:10.1016/j.advwatres.2014.06.008.

Zha, Y., T.-C. J. Yeh, W. A. Illman, T. Tanaka, P. Bruines, H. Onoe, and H. Saegusa (2015), What does hydraulic tomography tell us about fractured geological media? A field study and synthetic experiments, *J. Hydrol.*, *531*(1), 17–30, doi:10.1016/j.jhydrol.2015.06.013.

Zha, Y., T.-C. J. Yeh, W. A. Illman, T. Tanaka, P. Bruines, H. Onoe, H. Saegusa, D. Mao, S. Takeuchi, and J.-C. Wen (2016), An Application of Hydraulic Tomography to a Large-Scale Fractured Granite Site, Mizunami, Japan, *Groundwater*, *54*(6), 793–804, doi:10.1111/gwat.12421.

Zhao, Z., W. A. Illman, and S. J. Berg (2016), On the importance of geological data for hydraulic tomography analysis: Laboratory sandbox study. *J. Hydrol.*, *542*, 156–171, doi:10.1016/j.jhydrol.2016.08.061.

Zhao, Z., and W. A. Illman (2017), On the Importance of Geological Data for Three-dimensional Steady-State Hydraulic Tomography Analysis at a Highly Heterogeneous Aquifer-Aquitard System. *J. Hydrol.*, *544*, 640–657.

Zhang, J., and T.-C. J. Yeh (1997), An iterative geostatistical inverse method for steady flow in the vadose zone, *Water Resour. Res.*, *33*(1), 63–71, doi:10.1029/96WR02589.

Zhou, H., J. J. Gómez-Hernández, and L. Li (2014), Inverse methods in hydrogeology: Evolution and recent trends, *Adv. Water Resour.*, *63*, 22–37, doi:10.1016/j.advwatres.2013.10.014.

Zhu, J., and T. Yeh (2005), Characterization of aquifer heterogeneity using transient hydraulic tomography, *Water Resour. Res.*, *41*, W07028, doi:10.1029/2004WR003790.