RESEARCH ARTICLE

10.1002/2017WR021884

Key Points:

- The Reduced-Order Successive Linear Estimator can solve highly parameterized geostatistical inverse problem
- Approximation of covariance leads to affordable memory and computational cost
- The uncertainty of the estimates can be updated every iteration efficiently

Correspondence to:

L. Shi,
liangshs@whu.edu.cn

Citation:

Zha, Y., Yeh, T.-C. J., Illman, W. A., Zeng, W., Zhang, Y., Sun, F., et al. (2018). A reduced-order successive linear estimator for geostatistical inversion and its application in hydraulic tomography. *Water Resources Research*, 54, 1616–1632. <https://doi.org/10.1002/2017WR021884>

Received 18 SEP 2017

Accepted 10 FEB 2018

Accepted article online 16 FEB 2018

Published online 8 MAR 2018

A Reduced-Order Successive Linear Estimator for Geostatistical Inversion and its Application in Hydraulic Tomography

Yuanyuan Zha¹ , Tian-Chyi J. Yeh^{2,3} , Walter A. Illman⁴, Wenzhi Zeng¹ , Yonggen Zhang⁵ , Fangqiang Sun⁶, and Liangsheng Shi¹
¹State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan, China,

²Department of Hydrology and Atmospheric Sciences, University of Arizona, Tucson, Arizona, USA, ³Key Laboratory for Water Environment and Resources, Tianjin Normal University, Tianjin, China, ⁴Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, Ontario, Canada, ⁵Institute of Surface-Earth System Science, Tianjin University, Tianjin, China, ⁶Key Laboratory for Groundwater and Ecology in Arid and Semi-Arid Areas, China Geological Survey, Xi'an, China

Abstract Hydraulic tomography (HT) is a recently developed technology for characterizing high-resolution, site-specific heterogeneity using hydraulic data (n_d) from a series of cross-hole pumping tests. To properly account for the subsurface heterogeneity and to flexibly incorporate additional information, geostatistical inverse models, which permit a large number of spatially correlated unknowns (n_y), are frequently used to interpret the collected data. However, the memory storage requirements for the covariance of the unknowns ($n_y \times n_y$) in these models are prodigious for large-scale 3-D problems. Moreover, the sensitivity evaluation is often computationally intensive using traditional difference method (n_y forward runs). Although employment of the adjoint method can reduce the cost to n_d forward runs, the adjoint model requires intrusive coding effort. In order to resolve these issues, this paper presents a Reduced-Order Successive Linear Estimator (ROSLE) for analyzing HT data. This new estimator approximates the covariance of the unknowns using Karhunen-Loeve Expansion (KLE) truncated to n_{kl} order, and it calculates the directional sensitivities (in the directions of n_{kl} eigenvectors) to form the covariance and cross-covariance used in the Successive Linear Estimator (SLE). In addition, the covariance of unknowns is updated every iteration by updating the eigenvalues and eigenfunctions. The computational advantages of the proposed algorithm are demonstrated through numerical experiments and a 3-D transient HT analysis of data from a highly heterogeneous field site.

1. Introduction

Hydraulic tomography (HT) is an approach for characterizing the heterogeneity of spatially correlated hydraulic parameters of geologic media at high resolutions (Butler et al., 1999; Gottlieb & Dietrich, 1995; Yeh & Liu, 2000; Zhu & Yeh, 2005). Exploiting a series of man-made or naturally occurring hydraulic excitations (e.g., pumping tests, river-stage variations, or precipitations), HT collects a large number of drawdown-time data from different observation wells (Yeh et al., 2008). As this data set contains information about aquifer heterogeneity under different flow fields, HT requires an appropriate inverse model to derive the high-resolution spatial distribution of hydraulic parameters. The large number of data sets and a large number of parameters to be identified presents a significant challenge to the solution of inverse models. In order to overcome this challenge, classical approaches assign a small number of predetermined structures (e.g., zones or layers) for the aquifer to reduce the number of parameters to be identified or use pilot points. With a small number of predetermined structures, these approaches yield low-resolution and perhaps biased parameter fields (Yeh et al., 2015a, 2015b). Geostatistical approaches, by contrast, adopt a highly parameterized heterogeneous conceptual model, and estimate hydraulic parameters as well as their uncertainties at all elements or cells in the model domain (Kitanidis, 2015; Yeh & Liu, 2000). The parameter fields derived from these approaches yield unbiased predictions of flow and solute transport (Illman et al., 2012; Ni et al., 2009; Zha et al., 2017). Such benefits make geostatistical approaches attractive for HT analyses (Illman et al., 2015) and analyses in other fields (such as electrical resistivity tomography (Yeh et al., 2002), river-stage tomography (Yeh et al., 2009), data fusion of HT and tracer tomography (Yeh & Zhu, 2007), and fusion of environmental responses for basin-scale characterization (Yeh et al., 2008)).

Two geostatistical approaches have been widely applied to the field of hydraulic tomography: the Quasi-Linear Geostatistical Approach (QLGA) by Kitanidis (1995) and the Successive Linear Estimator (SLE) by Yeh et al. (1996), both of which have been used to interpret the hydraulic tomographic data from synthetic numerical experiments (e.g., Cardiff et al., 2013; Fienen et al., 2008; Yeh & Liu, 2000), laboratory sandboxes (Illman et al., 2007; Liu et al., 2007; Liu & Kitanidis, 2011; Zhao et al., 2015; Zhou & Cardiff, 2017; Zhou et al., 2016), and field tests (Cardiff et al., 2012; Illman et al., 2009; Hochstetler et al., 2015; Zha et al., 2016). The employment of adjoint approach in QLGA and SLE has been very useful in the HT applications since the cost for evaluating sensitivity matrix is reduced from the number of unknown parameters (n_y) to the number of selected fitting measurements (n_d). However, these two approaches may confront difficulty in 3-D, real-world HT applications due to two reasons. First, both approaches are computational demanding as the number (n_y) of unknown parameters becomes large since they require matrix multiplications and memory storage of the covariance matrix (with n_y^2 entries). Second, the inversion theoretically needs to include many noisy drawdown data to minimize effects of noise. Because of this, the adjoint method would require many simulations (n_d typically ranges from 10^2 to 10^4) and thus increase the computational burden of QLGA and SLE.

Considerable efforts have been devoted to circumventing this difficulty recently. For instance, Nowak et al. (2003) proposed an efficient algorithm to compute the cross-covariance of the unknowns and the observation data, and the autocovariance of the observations, using fast Fourier transform. Their method is most suitable for the cases when the covariance of the unknowns is Toeplitz (constant diagonals), which requires that the random field is omnidirectional, stationary and it is discretized on a regular equispaced grid. Liu and Kitanidis (2011) reformulated the Bayesian posterior distribution of the unknown parameters with a sparse representation. Due to the sparsity of the distribution, the inverse problem was solved efficiently. Subsequently, based on the reformulated posterior distribution, Liu et al. (2013) developed a geostatistical reduced-order model. In this model, the solution parameter space was mapped by the cross-covariance between parameters and observation data. The construction of the cross-covariance, however, required an evaluation of the sensitivity matrix. For reducing the computational cost of the sensitivity analysis, the sensitivity was evaluated only at the beginning of inversion or when the fitting error between the observation and simulated data cannot be improved. Later, Liu et al. (2014) developed a Krylov subspace method to avoid the direct solution of sensitivity matrix $\mathbf{J}$. Specifically, the products $\mathbf{J}\mathbf{x}$ and $\mathbf{J}^T\mathbf{x}$ are solved by the adjoint method at each iteration for any appropriately sized vector $\mathbf{x}$, and the forward model only needs to be evaluated twice.

Recently, Kitanidis and Lee (2014) proposed an efficient Principal Component Geostatistical Approach (PCGA) for large-dimensional inversion based on the Quasi-Linear Geostatistical Approach (QLGA). The randomized eigendecomposition approach was employed to approximate the prior covariance of parameters. With a so-called matrix-free method, the autocovariance of observations, as well as the cross-covariance of parameters and observations, were computed using the directional sensitivities, which are obtained from finite differencing outputs from the forward model. For each iteration, the required runs of the forward model can be equal to or less than the number of observation data. Also, the computational and storage cost of implementation of the inverse procedure scales roughly linearly with the number of unknown parameters (n_y). More importantly, the Jacobian matrix is not explicitly calculated, thus the proposed inverse approach can be easily adaptable to different forward models (e.g., flow, solute transport, or geophysics) and various data types. Otherwise, an adjoint-state model has to be coded for each forward model. Though the posterior covariance is not computed or stored in PCGA, the variances of the estimates were computed at the end of the iteration with matrix-vector products. Lee and Kitanidis (2014) subsequently applied this method to a large-scale HT with unknown parameters $n_y = 3 \times 10^6$ and a joint inversion problem of head and tracer test data. These applications demonstrated the advantages of the PCGA. More recently, a data reduction technique combined with the PCGA was developed by Lin et al. (2017) to improve the computational efficiency and to reduce the memory requirement. The success in improving the standard QLGA inspires us to develop an efficient method for SLE.

Although SLE and QLGA share similarities, a major difference between them is that the covariance of unknown parameters in SLE is updated to derive the residual covariance every iteration (Nowak & Cirpka, 2004; Zha et al., 2017). To be specific, the collected data are assimilated iteratively with a stochastic linear estimator in SLE to consider the nonlinear relationship between data and parameter as well as the

covariance of the parameters. The covariance of parameters initially specifies the spatial variability of the parameters. After conditioning the estimates of the parameters with data and the linear estimator, the covariance becomes the conditional covariance (or residual covariance: uncertainty associated with the estimates). Because this residual covariance is a linear approximation, it must be updated every iteration to reflect a further reduction in uncertainty of the estimates due to additional information extracted from the nonlinear relationship between data and parameters. That is, using this linearly approximated residual covariance, SLE then constructs new cross-covariance between parameter and data, and autocovariance of data (i.e., head) for the data assimilation at the next iteration. The continuous updating of residual covariance allows SLE to correctly address the uncertainty and enhance convergence of the inverse solution. However, updating the residual covariance of parameters is a major stumbling block for extending PCGA to SLE, since finding the eigenvalues and eigenvectors of a large residual covariance matrix is computationally expensive, if not prohibitive, for field-scale applications.

In this paper, we propose the use of truncated KLE to approximate the covariance of the unknown parameters such that the full-rank covariance matrix does not need to be explicitly computed and stored. Furthermore, the directional sensitivities (in the directions of n_{kl} eigenvectors) are calculated to form the covariance and cross-covariance matrix used in the cokriging-like inverse equation. The uncertainty of the estimates (conditional covariance) is then updated during each iteration via updating the leading eigenvalues and eigenfunctions associated with the residual covariance.

This paper is arranged as follows. We first revisit the Successive Linear Estimator (SLE) that forms the geostatistical inversion approach (section 2.1). We thereafter start from the truncated KLE to approximate the covariance of the unknown parameters, which has an analytical solution for a uniform grid (section 2.2). The Reduced-Order Successive Linear Estimator (ROSLE) is then developed with the help of directional sensitivity (section 2.3). In section 3, we test ROSLE with two steady state 2-D synthetic examples and a 3-D transient hydraulic tomography at a highly heterogeneous field site to demonstrate the superiority of the proposed method over SLE. Lastly, summary and conclusions are presented.

2. Methodology

2.1. Overview of the Successive Linear Estimator

Here we review the algorithm of the Successive Linear Estimator (SLE) used in HT analysis, which is originally proposed by Yeh et al. (1996) based on an iterative cokriging-like inverse model (Yeh et al., 1995). SLE adopts a highly parameterized conceptual model, which characterizes the unknown parameter as a spatial random field. As such, it discretizes the 3-D domain of the field site into N elements (material grid, which could be different from the forward grid used to solve the flow problem). Each element has a value for each hydraulic parameter (i.e., natural logarithms of hydraulic conductivity K [L/T] and specific storage S_s [1/T]— $\ln K$ and $\ln S_s$, respectively). Moreover, $\ln K$ and $\ln S_s$ can use different material grids ($N = N_k$ for $\ln K$ and N_s for $\ln S_s$, respectively) since $\ln S_s$ usually has lower spatial resolution based on HT data. The SLE then considers these hydraulic parameters as spatial stochastic processes (or random fields) with unconditional means $E[\ln K]$ and $E[\ln S_s]$ (where $E[-]$ denotes expectation), and the unconditional perturbations $\ln K - E[\ln K]$ and $\ln S_s - E[\ln S_s]$, respectively. These perturbations represent spatial variability of the parameters.

SLE estimates the most likely parameter value (i.e., conditional expectation or more appropriately conditional effective parameter value) for each element, conditioned with the observed drawdown data from the HT survey. It starts with some prior knowledge about mean values of the unknown parameter fields, which are combined as vector $\mathbf{Y}$ ($n_y \times 1$) in the numerical model. n_y is equal to N_k if only $\ln K$ is concerned, and it is $N_k + N_s$ if $\ln K$ and $\ln S_s$ are jointly estimated for transient problems.

Suppose during a pumping test, we have collected n_d observed heads in time and space, denoted by the data vector $\mathbf{d}^*$. The estimated parameter vector, given the observation, is $\mathbf{Y}_c$ (subscript c denotes conditional), and is iteratively determined using the following linear estimator:

$$\mathbf{Y}_c^{(r+1)} = \mathbf{Y}_c^{(r)} + \boldsymbol{\omega}^{(r)\top} (\mathbf{d}^* - \mathbf{d}^{(r)}) \quad (1)$$

where r is the iteration index; the vector $\mathbf{d}^{(r)}$ is the simulated heads at the observation locations and times obtained from the forward model, using the parameters obtained at iteration r . When $r = 0$, $\mathbf{Y}_c = E[\mathbf{Y}]$. The

coefficient matrix, ω ($n_d \times n_y$), denotes the weights, which assign the contribution of difference between the observed and simulated head at each observation location and time to a previously estimated parameter value at each element.

The coefficient matrix ω is determined by solving the following equations:

$$[\epsilon_{dd}^{(r)} + \theta^{(r)} \text{diag}(\epsilon_{dd}^{(r)})] \omega^{(r)} = \epsilon_{dy}^{(r)} \quad (2)$$

where ϵ_{dd} is the conditional (or residual) autocovariance of observation data and ϵ_{dy} is the residual cross-covariance between parameter and data. The parameter θ is a dynamic stability multiplier and $\text{diag}(\epsilon_{dd})$ is a stability matrix, which is a diagonal matrix with the same diagonal elements as ϵ_{dd} . The solution of equation (2) requires knowledge of autocovariance ϵ_{dd} and cross-covariance ϵ_{dy} , which can be derived from the first-order numerical approximation (Yeh & Liu, 2000):

$$\begin{aligned} \epsilon_{dd}^{(r)} &= \mathbf{J}_d^{(r)} \epsilon_{yy}^{(r)} \mathbf{J}_d^{(r)T}, \\ \epsilon_{dy}^{(r)} &= \mathbf{J}_d^{(r)} \epsilon_{yy}^{(r)} \end{aligned} \quad (3)$$

where $\mathbf{J}_d$ ($n_d \times n_y$) is the sensitivity (Jacobian) matrix of model predictions with respect to parameters estimated at current iteration. At iteration $r = 0$, the ϵ_{yy} ($n_y \times n_y$) is the unconditional covariance of parameters, which specifies the spatial variability of the parameter. It can be constructed from the assumed geologic structures of $\ln K$ and $\ln S_s$ (i.e., variance σ^2 , correlation scales, and covariance function). For $r \geq 1$, the residual or conditional (residual) covariance function is evaluated according to:

$$\epsilon_{yy}^{(r+1)} = \epsilon_{yy}^{(r)} - \omega^T \epsilon_{dy} \quad (4)$$

The above only describes the inverse algorithm using data from a single pumping test. In HT, data from several pumping tests are used conjunctively by adopting one of the two different approaches, e.g., Sequential SLE (SSLE) and Simultaneous SLE (SimSLE). Both SimSLE and SSLE are based on the Successive Linear Estimator (SLE) algorithm (Xiang et al., 2009; Zhu & Yeh, 2005).

A major drawback of SLE is that the covariance matrix requires a large memory storage. For instance, when $n_y = 10^5$, the storage for a double-precision covariance matrix ϵ_{yy} is approximately 150 GB, which is not yet commonly available for current personal computers. Previous applications of SLE for large-scale inverse problem utilized a reduced covariance matrix storage technique to save on memory (Zha et al., 2015, 2016). Specifically, when the cross-covariance value is smaller than a user-defined percentage of the variance, it will be considered as 0. The resultant banded matrix is stored at a much lower cost by a compressed row storage algorithm. However, the efficiency of this method could be low if the parameter field has long correlation scales. Furthermore, as an iterative solver, the computational costs for sensitivity matrix $\mathbf{J}_d$ and matrix-vector multiplications in equation (3) during each iteration are significant, albeit parallel computing approaches could be used. These difficulties are the major issue for applications of SLE to HT problems with a large number of unknowns (e.g., $\geq 10^5$).

2.2. Karhunen-Loeve Expansion

The KLE can be used to generate random fields based on given correlation functions (Ghanem & Spanos, 2003). In our case, the parameter random field $\mathbf{Y}$ can be expanded as (Zhang & Lu, 2004):

$$\mathbf{Y} = E[\mathbf{Y}] + \mathbf{g} \lambda^{1/2} \xi \quad (5)$$

where $E[\mathbf{Y}]$ is the mean of the discretized random field $\mathbf{Y}$, ξ ($n_y \times 1$) are a set of uncorrelated random variables with zero mean and unity variance, which means that $E[\xi_i] = 0$, $E[\xi_i \xi_j] = 1$, and $E[\xi_i \xi_j] = 0$, when $i \neq j$. Furthermore, $\mathbf{g}$ ($n_y \times n_y$) are orthogonal eigenfunctions, and λ ($n_y \times n_y$) is a diagonal matrix filling with corresponding nonnegative eigenvalues λ_i (where $i = 1, 2, \dots, n_y$, and they are rearranged in nonincreasing order). The eigenvalues and eigenvectors are obtained by eigenvalue decomposition of the covariance function of $\mathbf{Y}$:

$$\epsilon_{yy} = \mathbf{g}^T \lambda \mathbf{g} \quad (6)$$

A covariance matrix should be symmetrical based on the definition of the covariance. Thus, the eigendecomposition of the covariance matrix results in n_y real eigenvalues and mutually orthogonal eigenfunctions

that form a complete set. Although the standard eigendecomposition of an arbitrary covariance is time-consuming, we can take advantage of some fast decomposition approaches if the covariance fulfills some conditions. For instance, analytical solutions of $\mathbf{g}$ and λ are available for a separable exponential covariance function defined on a rectangular-shaped or brick-shaped domain (Ghanem & Spanos, 2003; Zhang & Lu, 2004). Moreover, the eigenvalues and eigenvectors of the circulant matrix (a special Toeplitz where each row vector is rotated one element to the right relative to the preceding row vector), which is obtained from circulant embedding of original covariance, can be solved efficiently using the fast Fourier transform algorithm (Nowak et al., 2003; Li & Cirpka, 2006). The obtained periodic eigenvectors defined on a regular grid can be easily mapped to the original domain. In the PCGA approach, the randomized eigendecomposition method (Halko et al., 2011) is employed to approximate the prior covariance. The computational costs for the latter two approaches are closely related to the cost of random field generation.

It should be noted that usually KLE is truncated according to the decaying of eigenvalues λ_i (where $i = 1, 2, \dots, n_y$). For instance, if the summation of the first n_{kl} eigenvalues dominates the summation of all eigenvalues, then only the first n_{kl} eigenvalues are kept in equation (6), and $\mathbf{g}$ in the eigenvalue decomposition is reduced to a $n_y \times n_{kl}$ matrix.

A norm of the normalized cumulated eigenvalues ($\beta = \sum_1^{n_{kl}} \lambda_i / \sum_1^{n_y} \lambda_i$) is usually employed to measure how well the truncated KLE approximates the original covariance. The n_{kl} eigenvalues are first summed up, i.e., $\sum_1^{n_{kl}} \lambda_i$, and then the summation is normalized by the summation of all eigenvalues ($\sum_1^{n_y} \lambda_i$), which is also equal to the variance times the area or volume of the domain (Lu & Zhang, 2004). Apparently, β is between 0 and 1, and it increases with the number of truncated terms (n_{kl}). A larger β indicates a better approximation accuracy of the truncated KLE.

2.3. Reduced-Order Successive Linear Estimator

Suppose for the first iteration ($r = 1$) we obtain the eigenvalues and eigenfunctions for the unconditional covariance functions. If we denote $\mathbf{f} = \mathbf{g}\lambda^{1/2}$ ($n_y \times n_{kl}$), the covariance and cross-covariance in equation (3) become:

$$\begin{aligned} e_{dd}^{(r)} &= \mathbf{J}_d^{(r)} \mathbf{f}^{(r)} \mathbf{f}^{(r)T} \mathbf{J}_d^{(r)T}, \\ e_{dy}^{(r)} &= \mathbf{J}_d^{(r)} \mathbf{f}^{(r)} \mathbf{f}^{(r)T} \end{aligned} \quad (7)$$

Since $\{\mathbf{g}_i\}$ or $\{\mathbf{f}_i\}$ with $i = 1, 2, \dots, n_{kl}$ form a complete set (Lu & Zhang, 2004), the coefficient matrix ω can also be spanned by $\mathbf{f}$

$$\omega^{(r)} = \mathbf{v}^{(r)} \mathbf{f}^{(r)T} = \mathbf{v}^{(r)} \lambda^{1/2} \mathbf{g}^T \quad (8)$$

where $\mathbf{v}$ ($n_d \times n_{kl}$) are the coefficients to form ω . Substituting equations (7) and (8) into equation (3), and we further denote $\mathbf{H} = \mathbf{J}_d \mathbf{f}$ ($n_d \times n_{kl}$), which is called directional gradient in Elsheikh et al. (2013), we have

$$[\mathbf{H}^{(r)} \mathbf{H}^{(r)T} + \theta^{(r)} \text{diag}(\mathbf{H}^{(r)} \mathbf{H}^{(r)T})] \mathbf{v}^{(r)} = \mathbf{H}^{(r)} \quad (9)$$

Again, the stabilization term $\theta^{(r)} \text{diag}(\mathbf{H}^{(r)} \mathbf{H}^{(r)T})$ is necessary to ensure a solution of this equation. After $\mathbf{v}$ is solved from equation (9), we can reconstruct ω by equation (8) and update the mean parameter field by equation (1):

$$\mathbf{Y}_c^{(r+1)} = \mathbf{Y}_c^{(r)} + \mathbf{f}^{(r)} \mathbf{v}^{(r)T} (\mathbf{d}^* - \mathbf{d}^{(r)}) \quad (10)$$

The key problem is to find $\mathbf{H}$. Elsheikh et al. (2013) and Kitanidis and Lee (2014) suggest using the first-order approximation to evaluate this matrix:

$$\begin{aligned} \mathbf{H}_i^{(r)} &= \mathbf{J}_d^{(r)} \mathbf{f}_i^{(r)} \approx \frac{1}{\delta_i} [G(\mathbf{Y}_c^{(r)} + \delta_i \mathbf{f}_i) - G(\mathbf{Y}_c^{(r)})] \\ i &= 1, 2, \dots, n_{kl} \end{aligned} \quad (11)$$

where $G()$ is the forward model that converts the hydraulic parameter $\mathbf{Y}$ ($n_y \times 1$) to simulated observation $\mathbf{d}$ ($n_d \times 1$). Apparently, $\mathbf{d}^{(r)}$ in equation (10) is equal to $G(\mathbf{Y}_c^{(r)})$. As noted by Kitanidis and Lee (2014) and Lee and Kitanidis (2014), the difference step δ_i is crucial to the successful calculation of $\mathbf{H}$. Since the vector $\mathbf{f}_i$

$(n_y \times 1)$ is scalable to $\sqrt{\lambda_i}$, it is favorable to set δ_i as a parameter inversely proportional to $\sqrt{\lambda_i}$, i.e., $\delta_i = \delta/\sqrt{\lambda_i}$, so that

$$\mathbf{H}_i^{(r)} \approx \frac{\sqrt{\lambda_i}}{\delta} [G(\mathbf{Y}_c^{(r)} + \delta \mathbf{g}_i) - G(\mathbf{Y}_c^{(r)})] \quad (12)$$

$$i=1, 2, \dots, n_{kl}$$

By using this approximation, only $n_{kl} + 1$ forward simulations are needed to construct $\mathbf{H}$. δ is a user-specified small number related to the machine round-off error (Kitanidis & Lee, 2014). The step size of finite difference method should not be too large (to avoid error introduced by approximating derivative using difference) nor too small (to avoid numerical instability due to catastrophic cancelation error). Using the groundwater flow model VSATF3 (Srivastava & Yeh, 1992; Zha et al., 2015), we find that it is preferable that the average relative change of element-wise parameter (in terms of K or S_s) is larger than 0.01% and smaller than 1%. Recall that $\mathbf{Y}$ is the natural logarithm of K or S_s , thus we have $0.01\% \leq E[\exp(\delta \mathbf{g}_i)] - 1 \leq 1\%$. Since $\mathbf{g}_i$ is obtained from eigenvalue decomposition, we have $\mathbf{g}_i^T \mathbf{g}_i = \|\mathbf{g}_i\|^2 = 1$. Thus, the average size of $\mathbf{g}_i$ component is $1/n_y$. Furthermore, notice that $E[\exp(\delta \mathbf{g}_i)] - 1 \approx E[\delta \mathbf{g}_i]$ when $E[\delta \mathbf{g}_i]$ is small, we can identify that the proper range of δ is between $10^{-4}/n_y$ and $10^{-2}/n_y$.

The sensitivity matrix $\mathbf{J}_d$ is calculated by direct finite difference method (used in PEST (Doherty, 2007) and UCODE (Poeter et al., 2005), $n_y + 1$ forward simulation runs) or adjoint method (used in QLGA and SLE, $n_d + 1$ forward simulation runs). Neither of these two methods are efficient since n_y and n_d are both very large in real-world 3-D HT applications. Thus, the finite difference method employed here has significant merits since it calculates the directional sensitivity matrix $\mathbf{H}$ instead of $\mathbf{J}_d$ with a cost of $n_{kl} + 1$ forward simulation runs.

After the first iteration, the conditional covariance can be decomposed in the same form

$$\boldsymbol{\varepsilon}_{ff}^{(r+1)} = \mathbf{f}^{(r+1)} \mathbf{f}^{(r+1)T} = \mathbf{g}^{(r+1)} \boldsymbol{\lambda}^{(r+1)} \mathbf{g}^{(r+1)T} \quad (13)$$

Starting from equation (4), the updating of the covariance matrix can be rewritten as

$$\begin{aligned} \boldsymbol{\varepsilon}_{ff}^{(r+1)} &= \boldsymbol{\varepsilon}_{ff}^{(r)} - (\mathbf{v}^{(r)} \mathbf{f}^{(r)T})^T \boldsymbol{\varepsilon}_{df}^{(r)} \\ &= \mathbf{f} \mathbf{f}^{(r)T} - \mathbf{f}^{(r)} \mathbf{v}^{(r)T} \mathbf{H}^{(r)} \mathbf{f}^{(r)T} \\ &= \mathbf{g}^{(r)} \boldsymbol{\lambda}^{(r)1/2} (\mathbf{I} - \mathbf{v}^{(r)T} \mathbf{H}^{(r)}) \boldsymbol{\lambda}^{(r)1/2T} \mathbf{g}^{(r)T} \end{aligned} \quad (14)$$

where $\mathbf{I}$ is a $n_{kl} \times n_{kl}$ identity matrix. Apparently, the conditional random field becomes nonstationary after the HT data are conditioned. In this situation, an analytical solution of the decomposition is not available anymore, while the direct solution of $\mathbf{f}^{(r+1)}$ is quite costly. Lu and Zhang (2004) proposed a method to obtain the eigenvalues and eigenvectors of the conditional covariance of kriging with manageable efforts. Inspired by their study, we propose an efficient method to obtain the eigenvalues and eigenvectors of the conditional covariance of SLE. First, we can span $\mathbf{g}^{(r+1)}$ based on $\mathbf{g}^{(r)}$:

$$\mathbf{g}^{(r+1)} = \mathbf{g}^{(r)} \mathbf{u}^{(r)} \quad (15)$$

where $\mathbf{u}$ is a $(n_{kl} \times n_{kl})$ matrix. Substituting equation (15) into equation (13), we have:

$$\boldsymbol{\varepsilon}_{ff}^{(r+1)} = \mathbf{g}^{(r)} \mathbf{u}^{(r)} \boldsymbol{\lambda}^{(r+1)} \mathbf{u}^{(r)T} \mathbf{g}^{(r)T} \quad (16)$$

By comparing equations (16) and (14), we have:

$$\boldsymbol{\lambda}^{(r)1/2} (\mathbf{I} - \mathbf{v}^{(r)T} \mathbf{H}^{(r)}) \boldsymbol{\lambda}^{(r)1/2T} = \mathbf{u}^{(r)} \boldsymbol{\lambda}^{(r+1)} \mathbf{u}^{(r)T} \quad (17)$$

Equation (17) describes an eigenvalue decomposition problem of a $(n_{kl} \times n_{kl})$ matrix $\mathbf{A}$, where $\mathbf{A} = \boldsymbol{\lambda}^{(r)1/2} (\mathbf{I} - \mathbf{v}^{(r)T} \mathbf{H}^{(r)}) \boldsymbol{\lambda}^{(r)1/2T}$. Since n_{kl} is usually not very large, we can update the conditional covariance matrix $\boldsymbol{\varepsilon}_{ff}^{(r+1)}$ using the updated eigenvalues $\boldsymbol{\lambda}^{(r+1)}$ and eigenvectors $\mathbf{g}^{(r+1)}$ via equation (15) with affordable cost.

2.4. Flowchart and Computation Cost of the ROSLE Algorithm

Figure 1 presents the flowchart of ROSLE algorithm. Initially, the inverse modeling starts with a prior mean and a covariance function with a variance and correlation scales. On a regular grid, the unconditional

Figure 1. Flowchart of ROSLE.

covariance $\varepsilon_{yy}^{(0)}$ is not explicitly calculated. Instead, using an analytical solution, we directly get λ and $\mathbf{g}$, $\mathbf{f} = \mathbf{g}\lambda^{1/2}$. During each iteration, we run the forward model $n_{kl} + 1$ times, one of which is given with mean value $E[\mathbf{Y}]$ and the rest are given with $E[\mathbf{Y}] + \delta\mathbf{g}_i$, where $i = 1, 2, \dots, n_{kl}$. Then, we construct matrix $\mathbf{H}$ ($n_d \times n_{kl}$) based on equation (12) and solve the cokriging-like equation (9) to get the coefficient $\mathbf{v}$. After that, we update the mean parameter $\mathbf{Y}^{(r+1)}$ based on equation (10), where $\mathbf{d}^{(r)}$ is obtained by running the forward model using conditional effective parameters, which is already obtained during the previous step. Finally, we update the eigenvalues and eigenvectors in order to proceed to next iteration. This final step involves several substeps. First, we conduct the eigenvalue decomposition of a symmetrical matrix $\mathbf{A}$ to get $\mathbf{u}$ and updated eigenvalues $\lambda^{(r+1)}$ based on equation (17), and then we construct the updated eigenvector using equation (15). The uncertainty at any iteration, if we are interested, can be constructed using equation (6). The iteration loop stops when certain criteria (e.g., reaching maximum iteration number, the head misfit error is small enough, or the parameter spatial variance is stable) are satisfied.

The detailed computation costs of ROSLE and standard SLE are listed in Table 1. Following the flowchart, SLE first evaluates the mean head using the guessed mean parameters by calling the forward model n_p (the number of pumping tests) times. Then, the sensitivity matrix $\mathbf{J}$ is evaluated by the adjoint model n_d (the number of observation) times. Both the forward and adjoint models are standard finite element models. Since the number of nodes and the efficiency of the sparse linear solver affect the computational cost of the finite element model, we

use the times of calling finite element models to indicate the major computational costs during the inversion. The cross-covariance and autocovariance in equation (3) are evaluated by matrix-vector multiplications. Subsequently, the cokriging coefficients in equation (2) are obtained by solving a linear system $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is ($n_d \times n_d$), and $\mathbf{x}$ and $\mathbf{b}$ are ($n_d \times n_y$). By solving the inverse of $\left[\varepsilon_{dd}^{(r)} + \theta^{(r)} \text{diag}(\varepsilon_{dd}^{(r)})\right]$, we can construct the coefficient by matrix product ($n_d \times n_d$) $\times$ ($n_d \times n_y$). Alternatively, some linear solver (e.g., dgefs in LINPACK (Dongarra et al., 1979)) can solve the first column of ω by factoring the left matrix $\left[\varepsilon_{dd}^{(r)} + \theta^{(r)} \text{diag}(\varepsilon_{dd}^{(r)})\right]$, and then the rest of the columns of ω can be obtained with much smaller cost. Finally, the mean and covariance of the parameter field are updated with matrix products ($n_y \times n_d$) $\times$ ($n_d \times 1$) and ($n_y \times n_d$) $\times$ ($n_d \times n_y$).

Table 1

Computational Costs of SLE and ROSLE at Every Iteration, Where n_p is the Number of Pumping Tests, n_y is the Number of Unknown Parameters, n_d is the Number of Observation Data, and n_{kl} is the Number of Truncated KLE Terms

	Equations	Cost
SLE		
Solution $\mathbf{d}^*$	$\mathbf{d} = \mathbf{G}(\mathbf{y})$	Call finite element solver $\times n_p$
Adjoint matrix $\mathbf{J}$	Adjoint equations	Call finite element solver $\times n_d$
Forming covariances	(3)	Matrix products: ($n_d \times n_y$) $\times$ ($n_y \times n_y$) + ($n_d \times n_y$) $\times$ ($n_y \times n_d$)
Cokriging coefficient	(2)	Solve linear system $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is ($n_d \times n_d$) and $\mathbf{x}$ and $\mathbf{b}$ are ($n_d \times n_y$)
Update of the parameters	(1)	Matrix products: ($n_y \times n_d$) $\times$ ($n_d \times 1$)
Update of the covariance	(4)	Matrix products: ($n_y \times n_d$) $\times$ ($n_d \times n_y$)
ROSLE		
Directional sensitivity matrix $\mathbf{H}$	(12)	Call finite element solver $n_p \times (n_{kl} + 1)$
Forming covariances	(7)	Matrix products: ($n_d \times n_{kl}$) $\times$ ($n_{kl} \times n_y$) + ($n_d \times n_{kl}$) $\times$ ($n_{kl} \times n_d$)
Cokriging coefficient $\mathbf{v}$	(9)	Solve linear system $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is ($n_d \times n_d$) and $\mathbf{x}$ and $\mathbf{b}$ are ($n_d \times n_{kl}$)
Update of the parameters	(10)	Matrix products: ($n_y \times n_{kl}$) $\times$ ($n_{kl} \times n_d$) $\times$ ($n_d \times 1$)
Obtain $\mathbf{f}$ and λ	(17)	Eigen decomposition of ($n_{kl} \times n_{kl}$) matrix + matrix products: ($n_{kl} \times n_d$) $\times$ ($n_d \times n_{kl}$) + ($n_{kl} \times n_{kl}$) $\times$ ($n_{kl} \times n_{kl}$)

In contrast, ROSLE first solves the directional sensitivity matrix $\mathbf{H}$ with a finite element solver $n_p \times (n_{kl} + 1)$ times. Then, the covariances are constructed with $\mathbf{H}$ by matrix products: $(n_d \times n_{kl}) \times (n_{kl} \times n_y) + (n_d \times n_{kl}) \times (n_{kl} \times n_d)$. In equation (9), we solve a smaller linear system $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is $(n_d \times n_d)$ and $\mathbf{x}$ and $\mathbf{b}$ are $(n_d \times n_{kl})$. The mean parameter field is updated with matrix products: $(n_y \times n_{kl}) \times (n_{kl} \times n_d) \times (n_d \times 1)$. In ROSLE, we update the eigenvalues and eigenvectors of the conditional covariance with matrix products $(n_{kl} \times n_d) \times (n_d \times n_{kl}) + (n_{kl} \times n_{kl}) \times (n_{kl} \times n_{kl})$ and eigenvalue decomposition of matrix $(n_{kl} \times n_{kl})$.

In ROSLE, the eigenvalues and eigenvectors of the unconditional covariance are obtained from analytical solutions or fast Fourier transform before the first iteration. Moreover, the conditional covariance can be constructed at any iteration by eigenvalues and corresponding eigenvectors with matrix products $(n_y \times n_{kl}) \times (n_{kl} \times n_y)$. These efforts not listed in Table 1 are not required for every iteration.

The maximum storage requirement for SLE is $\varepsilon_{yy} (n_y \times n_y)$. On the other hand, ROSLE only requires $(n_{kl} \times n_y + n_{kl})$ for the storage of parameter covariance information.

3. Tested Examples

In this section, we consider two HT cases to examine the performance of the proposed ROSLE algorithm by comparing it with the results from original SLE. Case 1 involves synthetic aquifers, and case 2 is an application to field HT experiments. The code was executed using eight cores (Intel Xeon E5-2630 v3 CPUs running at 2.4 GHz with 96 GB of RAM) on a cluster at the Super Computer Center at Wuhan University (<http://hpc.whu.edu.cn/>). During the forward/adjoint modeling and the matrix-vector multiplication, MPI (Gropp et al., 1999) is used to distribute the workload to different cores.

3.1. Case 1

Case 1 considers steady state hydraulic tomographic surveys in two 2-D synthetic aquifers with different types of heterogeneity in K fields. In scenario 1, the aquifer $\ln K$ heterogeneity is represented by an exponential covariance function, while the other (scenario 2) is a Gaussian covariance function. Applications to these two different covariance functions aim to demonstrate the effects of the number of terms utilized in KLE.

The synthetic aquifer has dimensions of 800 m in length and 400 m in height and is discretized into 800×600 elements. Each element is assigned a random $\ln K$ value, assuming it follows a jointly normal distribution with mean $\mu = 0$, variance $\sigma^2 = 1.0$, and the given covariance functions. The correlation scales are 200 and 50 m in x and y directions, respectively, for these two scenarios. The “true” K fields for scenario 1 and scenario 2 are given in Figures 2a and 2b, respectively. According to these two figures, the K field resulting from the exponential covariance function is more irregular than the other.

We impose constant-head boundary conditions on the left and right, and no-flow boundary conditions at the top and bottom of the aquifer. Twelve pumping tests (with unit pumping rate) are simulated in these aquifers at 12 locations (squares), and pressure responses at steady state conditions are collected at other locations (circles) in Figures 2a and 2b, excluding the pumping locations. This simulation leads to $77 \times 12 = 924$ hydraulic data for each of the two scenarios.

Figure 2. The true $\ln K$ fields in case 1 (left: scenario 1 using exponential covariance function and right: scenario 2 using Gaussian covariance function).

Afterward, these hydraulic data, aquifer geometry, boundary conditions, and the mean, variance, correlation scales, and covariance functions of the true K fields) are input to both SLE and ROSLE to reconstruct the “true” value in each of the 480,000 elements. However, the number of elements ($n_y = 480,000$) leads to a covariance of approximately 1,716 GB (assuming double precision) in SLE, which is too large to be stored in our computing equipment. Instead, we employ SLE to estimate the $\ln K$ value at a coarse inverse grid (160×80 , with 5×5 element size, leading to a smaller covariance storage of 2.75 GB) in each scenario (see Table 2 for details). In contrast, ROSLE (using a fine grid with $n_y = 480,000$) requires approximately 0.71 GB storage for eigenvectors with $n_{kl} = 200$. ROSLE requires calling the forward model 192, 612, or 2,412 times (corresponds to $n_{kl} = 15, 50$, or 200), while SLE needs to call forward and adjoint models 936 times (Table 2).

The estimated $\ln K$ fields for scenario 1 (exponential covariance) and scenario 2 (Gaussian covariance) are displayed in the left column and the right column of Figure 3, respectively. Results from ROSLE using 15, 50, and 200 terms in KLE are illustrated in rows 1, 2, and 3, respectively. The last row displays the estimates using SLE with coarse grids for the two scenarios. The comparison between estimated and true K for Figures 3a, 3b, 3c, and 3d are illustrated in Figures 4a, 4b, 4c, and 4d. The scatter plots for scenario 2 are displayed in Figures 4e, 4f, 4g, and 4h.

Comparing the results in the left column of Figure 3 (exponential covariance) with the “true” field in Figure 2a, we observe that ROSLE is able to recover the major high K and low K zones with only 15 truncated terms. By retaining more truncated terms (50 or 200) in ROSLE, some small-scale features appear. However, as indicated in the scatterplots (Figures 4a–4c), these small-scale features do not lead to the improvements of the statistical norms, namely L1 (mean absolute error), L2 (mean squared error), as well as the R^2 (coefficient of determination), or the slope and intercept of the fitted linear relationship between the estimates and the true values. On the other hand, SLE can reveal the large-scale features using a coarse inverse grid. The scatter plot for SLE is shown in Figure 4d. Note that in this plot, the estimated K value in each coarse grid of SLE is plotted against the averaged true K value of 25 fine grids around the center of the coarse grid. This averaging may explain the better performance of SLE in comparison to those from ROSLE.

Regarding scenario 2, where the covariance is Gaussian, the estimates from ROSLE using $n_{kl} = 15$ (Figure 3e) capture the large-scale features, but the shapes are distorted, as we compare it with the true field in Figure 2b. This poor performance is also reflected in the scatter plot (Figure 4e). However, the performance of ROSLE is significantly improved (e.g., R^2 increases from 0.59 to 0.91) when we increase n_{kl} to 50 (Figures 3f and 4f). The norms are further improved if n_{kl} is increased to 200 (Figures 3g and 4g).

The different behaviors of scenarios 1 and 2 in case 1 can be explained by Figure 5, which presents the norm of the normalized cumulated eigenvalues (β) as a function of the number of eigenvalues (n_{kl}) used in the truncated KLE. As demonstrated in Figure 5, the red curve (exponential covariance in scenario 1) shows a quick increase at the first few leading eigenvalues, while the increase of the green curve (Gaussian covariance in scenario 2) is slow. This explains the fact that with $n_{kl} = 15$, the eigenvalues and vectors are not accurate in representing the true covariance, leading to inferior results. On the other hand, the green curve reaches one after $n_{kl} \approx 200$. This is because the random field generated by Gaussian covariance does not

Table 2

Forward (Adjoint) Runs, CPU Time (Using Eight Cores), and Storage of the Parameter Covariance for the Numerical Cases

Case	Model	Forward (adjoint) runs	CPU time (h)	Memory storage (GB)
Case 1 ($n_y = 19,200$)	SLE	$924 + 12 = 936$	0.78 ^a	2.75
	ROSLE, $n_{kl} = 50$	$50 \times 12 + 12 = 612$	0.35 (−55.1%) ^b	0.01 (−99.6%)
Case 1 ($n_y = 480,000$)	SLE	$924 + 12 = 936$	NA ^c	1,716
	ROSLE, $n_{kl} = 15$	$15 \times 12 + 12 = 192$	0.35	0.05 (−99.9%)
	ROSLE, $n_{kl} = 50$	$50 \times 12 + 12 = 612$	0.96	0.18 (−99.9%)
	ROSLE, $n_{kl} = 200$	$200 \times 12 + 12 = 2,412$	3.02	0.71 (−99.9%)
Case 2 ($n_y = 23,328$)	SLE	$696 + 4 = 700$	4.12	8.11
	ROSLE, $n_{kl} = 50$	$50 \times 4 + 4 = 204$	1.08 (−73.8%)	0.02 (−99.8%)
	ROSLE, $n_{kl} = 300$	$300 \times 4 + 4 = 1,204$	4.96 (+20.4%)	0.10 (−98.8%)

^aAveraged for scenarios 1 and 2 in case 1. ^bA relative increase (+) or saving (−) for the computational cost of ROSLE compared to that of SLE. ^cNot available due to prohibitively large storage requirement.

Figure 3. The estimated $\ln K$ fields in case 1 (left: scenario 1 with exponential covariance function and right: scenario 2 with Gaussian covariance function).

show small-scale features (Figure 2b), and thus ROSLE in scenario 2 with $n_{kl} = 50$ or 200 yields good estimates. Since the true field of scenario 1 (Figure 2a) contains small-scale features, β asymptotically approaches one slowly. For instance, the truncated KLE only represents 92% of the total summed eigenvalues even $n_{kl} = 1,000$. This indicates that truncated method is prone to a loss of accuracy in the approximation of exponential covariance function. However, this does not necessarily lead to the deterioration of the inversion results. It is a well-known fact that the resolution of HT or inverse models is dictated by the information content of observations (dense, high-quality observations lead to high-resolution estimates). The insignificant improvement of scenario 1 with increasing n_{kl} indicates that the current well network is not able to accurately depict the small-scale features.

In the residual variance plot (Figure 6), the n_{kl} eigenvalues are multiplied by $1/\beta$ to ensure consistency and the unconditional covariance expressed by truncated eigenvalues has the same variance as the original one. Similar to the results of the estimated $\ln K$ fields, an increase of n_{kl} leads to better residual variance estimation. It is also noticeable (Figures 6a and 6b) that the ROSLE may severely underestimate the residual variance if n_{kl} is too small.

Figure 7 presents the data-fitting performance of case 1. Overall, the bottom scatterplots have better performances except for the leftmost one (Figure 7e). This is reasonable because the random field in scenario 2 has only large-scale features, and the inverse model can reduce the head misfit error if n_{kl} is large enough. On the other hand, an increase of n_{kl} does not lead to significant improvement of data-fitting performance

Figure 4. The inverted versus true $\ln K$ results (arithmetic averages over the 5×5 coarse grids are taken for the true $\ln K$ and those from ROSLE to facilitate comparison with SLE) for case 1 (top: scenario 1 with exponential covariance function and bottom: scenario 2 with Gaussian covariance function). The color indicates relative scatter point density.

in scenario 1 (Figures 7a–7d). The data-fitting performance of SLE (Figure 7d) suggests that SLE, although using a coarse inverse grid, has better performance than ROSLE ($L2$ of 1.3 versus values more than 3.0).

3.2. Case 2

In case 2, ROSLE is applied to a hydraulic tomographic survey conducted at the North Campus Research Site (NCRS) in Waterloo, Ontario, Canada (Berg & Illman, 2011). It consists of four continuous multichannel tubing (CMT) wells containing seven channels as observation ports, and five multilevel pumping wells (PW) containing three to five channels (Figure 8). For the CMT wells, the screens are spaced 2 m apart with the

uppermost screens located between 4.5 and 5.5 m below ground surface (mbgs), and the deepest screens are 16.5–17.5 mbgs. The simulation domain is $45 \text{ m} \times 45 \text{ m} \times 18 \text{ m}$ (Figure 8), and the dimensions of each element are $2.5 \text{ m} \times 2.5 \text{ m} \times 0.25 \text{ m}$, with a total of 26,353 nodes and 23,328 elements. The prior parameter fields are described by a uniform mean $K = 0.1 \text{ m/d}$, mean $S_s = 10^{-4} \text{ 1/m}$, variances of $\sigma^2_{\ln K} = 5.0$ and $\sigma^2_{\ln S_s} = 2.0$, horizontal isotropic correlation scales of 10 m and a vertical correlation scale of 2.5 m. The boundary conditions are no flow for top and bottom and constant head for the four lateral boundaries. In addition, the head is assumed hydrostatic before each pumping test.

A total of four pumping tests (with pumping taking place at intervals PW4–3, PW1–3, PW5–3, and PW3–3) conducted at the NCRS are included in the inversion. The locations, durations, and the rates of the tests can be found in Berg and Illman (2011). During each pumping test, 44–48 drawdown curves are collected from different intervals. For each drawdown-time curve, three to four data are sampled according to the strategy proposed by Sun et al. (2013). There are a total of 696 data points. We perform the joint inversion of $\ln K$ and $\ln S_s$ using ROSLE ($n_{kl} = 50,300$) and SLE. A reduced covariance matrix storage technique is used in SLE to save memory. Specifically, the entity of covariance matrix will be set to zero if it is smaller than a user-

Figure 5. The norm of the normalized cumulated eigenvalues (β) as a function of the number of eigenvalues (n_{kl}) for case 1 and case 2. Note that the eigenvalues are rearranged in nonincreasing order.

Figure 6. The residual variance of estimated $\ln K$ fields in case 1 (left: scenario 1 and right: scenario 2).

defined value ($0.05\sigma^2$). The resultant banded matrix then is stored at a lower cost by compressed row storage.

Figure 9 presents the estimated $\ln K$ and the variance of the estimates for case 2 ($\ln S_s$ estimates are not shown). All the estimated $\ln K$ values exhibit strong layering with alternative high K and low K layers. This is consistent with previous studies (Berg & Illman, 2011). More importantly, the estimated results using ROSLE are very similar to that from SLE. Similarly, the residual variance maps indicate that the uncertainty is smaller inside the well field, especially at the middle layer. It is also visible that the residual variance estimations from ROSLE show artificial layers along the domain boundary. This is possibly due to that the correlation scale in the z -direction is very small, and in order to capture the true variability, ROSLE requires a large number of eigenvalues. As shown in Figure 5, the β values are only 0.3 and 0.6 for the inversions using $n_{kl} = 50$ and 300, respectively.

As shown in Figure 10, ROSLE ($n_{kl} = 50, 300$) and SLE exhibit similar data-fitting performances, although there is a significant loss of prior covariance using the truncated KLE in this case. Since the estimated $\ln K$ from the three inversions are very similar, we may conclude that the accuracy of parameter covariance has less impact on the inversion in case 2 than that in case 1. A possible reason may be that for real 3-D field case, the data collected from HT test are sparse and less informative (due to noise), compared to those sampled from the synthetic case. Using a small number of n_{kl} in ROSLE, although losing accuracy, it still preserves the major structural features of the aquifer (e.g., the strong anisotropy in case 2) and leads to

Figure 7. The data-fitting performances for case 1 (top: scenario 1 and bottom: scenario 2).

satisfactory estimates. In case 2, the full storage of covariances of $\ln K$ and $\ln S_s$ requires approximately 8 GB (Table 2), while ROSLE using $n_{kl} = 50$ and 300 only requires 18 MB (saving 99.8%) and 108 MB (saving 98.7%). During each iteration, SLE requires calling forward or adjoint model 700 ($n_d + n_p$) times while ROSLE requires calling forward model 204 ($n_p n_{kl} + n_p$) or 1,204 times (Table 2).

Since case 2 is a real-world example, the assumption about the prior covariance has uncertainty. Due to the Bayesian nature of SLE or ROSLE, the prior covariance is important when the observation data are not enough to obtain the heterogeneity at the desired resolution. The residual variance map (e.g., Figures 9d, 9e, and 9f) can be used to assess the uncertainty of the estimates as well as the impact of the uncertain prior information. The prior covariance is also important to determine a proper number of truncated KLE terms (n_{kl}) based on the value of β .

Figure 8. The plan view and the well network of case 2 (Tso et al., 2016). (a) Plan view: the enclosed area (15 m $\times$ 15 m) is bounded by pumping and observations wells and (b) the well locations.

Figure 9. The inversion results for case 2 (left: inverted $\ln K$ and right: uncertainty of the estimates).

Figure 10. The data-fitting performance of case 2.

4. Summary and Conclusions

In this paper, we have developed a Reduced-Order Successive Linear Estimator (ROSLE) for large-scale geostatistical inversion, and hydraulic tomography data interpretation has been used to test the developed method. We use the truncated KLE to approximate the prior covariance function, which has an analytical solution for the uniform grid. As the linear estimator in SLE only requires the autocovariance of the observation and the cross-covariance between parameters and observations, the computation of element-wise sensitivity could be avoided. That is, ROSLE calculates the directional sensitivities, in the directions of the eigenvectors of the parameter covariance, by a finite difference approach with several forward runs. The conditional covariance is then updated recursively during each iteration by updating the eigenvalue and eigenvectors via the solution of a small eigenvalue decomposition problem.

We test the proposed algorithm with synthetic data and field data collected at a highly heterogeneous field site. The inverse solutions from ROSLE are very close to those using SLE. In the meantime, the computational cost is similar to SLE but the storage requirement are significantly less. The numerical test demonstrates that analyzing large-scale 3-D hydraulic tomography data with the number of elements up to 10^6 is not formidable and can be accomplished at a manageable computational cost. According to the tested examples, ROSLE is efficient when the prior unknown parameter has large correlation scale. In this case, only a few truncated KLE terms can properly approximate the prior covariance with much less computation cost compared to SLE. When the prior parameter has an insignificant spatial correlation structure (short correlation scale), a direct compressed storage of the prior covariance by setting some small entities of the covariance matrix as zeros (to solve the memory storage problem) may be more efficient than ROSLE. ROSLE require calling the forward model $n_{kl} + 1$ times, while SLE requires calling the forward/adjoint model $n_d + 1$ times. Independence of the cost of ROSLE on observation number n_d is a promising feature since high-resolution geostatistical inversion for real-world problems requires a large number of noisy observations.

Lastly, ROSLE is also very attractive for different types of geostatistical inverse problems in other science fields because the inverse algorithm now is independent of solving the adjoint-state equation associated with problem-specific governing equations. This is particularly useful and convenient for assimilating many different types of information (such as hydraulic head, tracer, electric potential, geophysical attributes as well as responses from other environmental processes). Furthermore, the size of covariance and cross-covariance between different processes could be reduced significantly. For this reason, the simultaneous assimilation of all data from different measurements now become possible and appropriately addresses conditional uncertainty associated with the results. These features are attractive for the applications of ROSLE to many other science and engineering fields. Moreover, since the residual covariance information can be stored in output files in terms of eigenvalues and eigenvectors, it can be retrieved for sequential Bayesian inversion if more data are collected in the future. For instance, the residual covariance of $\ln K$ field can be used for conditional realizations of solute transport simulation, or prior information of additional geophysical inversions.

Acknowledgments

The present study was supported by the National Natural Science Foundation of China (grants 51779179, 51609173, 51479144, and 51522904). The second author acknowledges funding by CRDF under the award (DAA2-15-61224-1): hydraulic tomography in shallow alluvial sediments: Nile River Valley, Egypt, and the Global Expert award through Tianjin Normal University from the Thousand Talents Plan of Tianjin City. F. Sun would like to acknowledge the Special Fund for Public Industry Research from Ministry of Land and Resources of China (No. 201511047). The authors are grateful for the thoughtful review and suggestions by Alberto Bellin (Associate Editor), Kris Kuhlman, and two anonymous reviewers. The program and the data used in this study can be downloaded from <https://doi.org/10.13140/RG.2.2.11775.71845>.

References

- Berg, S. J., & Illman, W. A. (2011). Three-dimensional transient hydraulic tomography in a highly heterogeneous glaciofluvial aquifer-aquitard system. *Water Resources Research*, 47, W10507. <https://doi.org/10.1029/2011WR010616>
- Butler, J. J., McElwee, C. D., & Bohling, G. C. (1999). Pumping tests in networks of multilevel sampling wells: Motivation and methodology for measurement. *Water Resources Research*, 35(11), 3553–3560. <https://doi.org/10.1029/1999WR900231>
- Cardiff, M., Bakhos, T., Kitanidis, P., & Barrash, W. (2013). Aquifer heterogeneity characterization with oscillatory pumping: Sensitivity analysis and imaging potential. *Water Resources Research*, 49, 5395–5410. <https://doi.org/10.1002/wrcr.20356>
- Cardiff, M., Barrash, W., & Kitanidis, P. K. (2012). A field proof-of-concept of aquifer imaging using 3-D transient hydraulic tomography with modular, temporarily-emplaced equipment. *Water Resources Research*, 48, W05531. <https://doi.org/10.1029/2011WR011704>
- Doherty, J. (2007). *PEST: Model-independent parameter estimation user manual* (5th ed.). Brisbane, Australia: Watermark Numerical Computing.
- Dongarra, J. J., Bunch, J., Moler, C., & Stewart, G. W. (1979). *LINPACK user's guide*. Philadelphia, PA: SIAM.
- Elsheikh, A. H., Wheeler, M. F., & Hoteit, I. (2013). An iterative stochastic ensemble method for parameter estimation of subsurface flow models. *Journal of Computational Physics*, 242, 696–714. <https://doi.org/10.1016/j.jcp.2013.01.047>
- Fienen, M. N., Clemo, T., & Kitanidis, P. K. (2008). An interactive Bayesian geostatistical inverse protocol for hydraulic tomography. *Water Resources Research*, 44, W00B01. <https://doi.org/10.1029/2007WR006730>
- Ghanem, R. G., & Spanos, P. D. (2003). *Stochastic finite elements: A spectral approach*. North Chelmsford, MA: Courier Corporation.
- Gottlieb, J., & Dietrich, P. (1995). Identification of the permeability distribution in soil by hydraulic tomography. *Inverse Problems*, 11, 353–360.
- Gropp, W., Lusk, E., & Skjellum, A. (1999). *Using MPI: Portable parallel programming with the message passing interface* (2nd ed.). Cambridge, MA: MIT Press.
- Halko, N., Martinsson, P. G., & Tropp, J. A. (2011). Finding structure with randomness: Probabilistic algorithms for matrix decompositions. *SIAM Review*, 53(2), 217–288.
- Hochstetler, D. L., Barrash, W., Leven, C., Cardiff, M., Chidichimo, F., & Kitanidis, P. K. (2015). Hydraulic tomography: Continuity and discontinuity of high-K and low-K zones. *Ground Water*, 54(2), 1–15. <https://doi.org/10.1111/gwat.12344>
- Illman, W. A., Berg, S. J., & Yeh, T.-C. J. (2012). Comparison of approaches for predicting solute transport: Sandbox experiments. *Ground Water*, 50(3), 421–431. <https://doi.org/10.1111/j.1745-6584.2011.00859.x>
- Illman, W. A., Berg, S. J., & Zhao, Z. (2015). Should hydraulic tomography data be interpreted using geostatistical inverse modeling? A laboratory sandbox investigation. *Water Resources Research*, 51, 3219–3237. <https://doi.org/10.1002/2014WR016552>
- Illman, W. A., Liu, X., & Craig, A. (2007). Steady-state hydraulic tomography in a laboratory aquifer with deterministic heterogeneity: Multi-method and multiscale validation of hydraulic conductivity tomograms. *Journal of Hydrology*, 341(3–4), 222–234. <https://doi.org/10.1016/j.jhydrol.2007.05.011>
- Illman, W. A., Liu, X., Takeuchi, S., Yeh, T.-C. J., Ando, K., & Saegusa, H. (2009). Hydraulic tomography in fractured granite: Mizunami Underground Research site, Japan. *Water Resources Research*, 45, W01406. <https://doi.org/10.1029/2007WR006715>
- Kitanidis, P. K. (1995). Quasi-linear geostatistical theory for inversing. *Water Resources Research*, 31(10), 2411–2419. <https://doi.org/10.1029/95WR01945>
- Kitanidis, P. K. (2015). Persistent questions of heterogeneity, uncertainty, and scale in subsurface flow and transport. *Water Resources Research*, 51, 5888–5904. <https://doi.org/10.1002/2015WR017639>
- Kitanidis, P. K., & Lee, J. (2014). Principal Component Geostatistical Approach for large-dimensional inverse problems. *Water Resources Research*, 50, 5410–5427. <https://doi.org/10.1002/2014WR015483>
- Lee, J., & Kitanidis, P. (2014). Large-scale hydraulic tomography and joint inversion of head and tracer data using the Principal Component Geostatistical Approach (PCGA). *Water Resources Research*, 50, 5410–5427. <https://doi.org/10.1002/2014WR015483>
- Li, W., & Cirpka, O. A. (2006). Efficient geostatistical inverse methods for structured and unstructured grids. *Water Resources Research*, 42, W06402. <https://doi.org/10.1029/2005WR004668>
- Lin, Y., Le, E. B., O'malley, D., Vesselinov, V. V., & Bui-Thanh, T. (2017). Large-scale inverse model analyses employing fast randomized data reduction. *Water Resources Research*, 53, 6784–6801. <https://doi.org/10.1002/2016WR020299>
- Liu, X., Illman, W. A., Craig, A. J., Zhu, J., & Yeh, T.-C. J. (2007). Laboratory sandbox validation of transient hydraulic tomography. *Water Resources Research*, 43, W05404. <https://doi.org/10.1029/2006WR005144>
- Liu, X., & Kitanidis, P. K. (2011). Large-scale inverse modeling with an application in hydraulic tomography. *Water Resources Research*, 47, W02501. <https://doi.org/10.1029/2010WR009144>
- Liu, X., Zhou, Q., Birkholzer, J., & Illman, W. A. (2013). Geostatistical reduced-order models in underdetermined inverse problems. *Water Resources Research*, 49, 6587–6600. <https://doi.org/10.1002/wrcr.20489>
- Liu, X., Zhou, Q., Kitanidis, P. K., & Birkholzer, J. T. (2014). Fast iterative implementation of large-scale nonlinear geostatistical inverse modeling. *Water Resources Research*, 50, 198–207. <https://doi.org/10.1002/2012WR013241>
- Lu, Z., & Zhang, D. (2004). Conditional simulations of flow in randomly heterogeneous porous media using a KL-based moment-equation approach. *Advances in Water Resources*, 27(9), 859–874. <https://doi.org/10.1016/j.advwatres.2004.08.001>
- Ni, C., Yeh, T., & Chen, J. (2009). Cost-effective hydraulic tomography surveys for predicting flow and transport in heterogeneous aquifers. *Environmental Science & Technology*, 43(10), 3720–3727.
- Nowak, W., & Cirpka, O. A. (2004). A modified Levenberg-Marquardt algorithm for quasi-linear geostatistical inversing. *Advances in Water Resources*, 27(7), 737–750. <https://doi.org/10.1016/j.advwatres.2004.03.004>
- Nowak, W., Tenklev, S., & Cirpka, O. A. (2003). Efficient computation of linearized cross-covariance and auto-covariance matrices of interdependent quantities. *Mathematical Geosciences*, 35(1), 53–66. <https://doi.org/10.1023/A:1022365112368>
- Poeter, E. P., Hill, M. C., Banta, E., Mehl, R. S., & Christensen, S. (2005). *UCODE_2005 and six other computer codes for universal sensitivity analysis, calibration, and uncertainty evaluation* (USGS Techniques and Methods 6-A11). Reston, VA: United States Geological Survey.
- Srivastava, R., & Yeh, T.-C. J. (1992). A three-dimensional numerical model for water flow and transport of chemically reactive solute through porous media under variably saturated conditions. *Advances in Water Resources*, 15(5), 275–287. [https://doi.org/10.1016/0309-1708\(92\)90014-5](https://doi.org/10.1016/0309-1708(92)90014-5)
- Sun, R., Yeh, T.-C. J., Mao, D., Jin, M., Lu, W., & Hao, Y. (2013). A temporal sampling strategy for hydraulic tomography analysis. *Water Resources Research*, 49, 3881–3896. <https://doi.org/10.1002/wrcr.20337>
- Tso, C.-H. M., Zha, Y., Yeh, T.-C. J., & Wen, J.-C. (2016). The relative importance of head, flux, and prior information in hydraulic tomography analysis. *Water Resources Research*, 52, 3–20. <https://doi.org/10.1002/2015WR017191>

- Xiang, J., Yeh, T.-C. J., Lee, C.-H., Hsu, K.-C., & Wen, J.-C. (2009). A simultaneous successive linear estimator and a guide for hydraulic tomography analysis. *Water Resources Research*, 45, W02432. <https://doi.org/10.1029/2008WR007180>
- Yeh, T.-C. J., Gutjahr, A. L., & Jin, M. (1995). An iterative cokriging-like technique for groundwater flow modeling. *Ground Water*, 33(1), 33–41.
- Yeh, T.-C. J., Khaleel, R., & Carroll, K. C. (2015a). *Flow through heterogeneous geologic media*. Cambridge, UK: Cambridge University Press.
- Yeh, T.-C. J., Jin, M., & Hanna, S. (1996). An iterative stochastic inverse method: Conditional effective transmissivity and hydraulic head fields. *Water Resources Research*, 32(1), 85–92. <https://doi.org/10.1029/95WR02869>
- Yeh, T.-C. J., Lee, C.-H., Hsu, K.-C., Illman, W. A., Barrash, W., Cai, X., et al. (2008). A view toward the future of subsurface characterization: CAT scanning groundwater basins. *Water Resources Research*, 44, W03301. <https://doi.org/10.1029/2007WR006375>
- Yeh, T.-C. J., & Liu, S. (2000). Hydraulic tomography: Development of a new aquifer test method. *Water Resources Research*, 36(8), 2095–2105. <https://doi.org/10.1029/2000WR900114>
- Yeh, T.-C. J., Liu, S., Glass, R. J., Baker, K., Brainard, J. R., Alumbaugh, D., et al. (2002). A geostatistically based inverse model for electrical resistivity surveys and its applications to vadose zone hydrology. *Water Resources Research*, 38(12), 1278. <https://doi.org/10.1029/2001WR001204>
- Yeh, T.-C. J., Xiang, J., Suribhatla, R. M., Hsu, K.-C., Lee, C.-H., & Wen, J.-C. (2009). River stage tomography: A new approach for characterizing groundwater basins. *Water Resources Research*, 45, W05409. <https://doi.org/10.1029/2008WR007233>
- Yeh, T.-C. J., & Zhu, J. (2007). Hydraulic/partitioning tracer tomography for characterization of dense nonaqueous phase liquid source zones. *Water Resources Research*, 43, W06435. <https://doi.org/10.1029/2006WR004877>
- Yeh, T. J., Mao, D., Zha, Y., Wen, J., Wan, L., Hsu, K., et al. (2015b). Uniqueness, scale, and resolution issues in groundwater model parameter identification. *Water Science and Engineering*, 8(3), 175–194. <https://doi.org/10.1016/j.wse.2015.08.002>
- Zha, Y., Yeh, T. J., Illman, W. A., Onoe, H., Mok, C. M. W., Wen, J.-C., et al. (2017). Incorporating geologic information into hydraulic tomography: A general framework based on geostatistical approach. *Water Resources Research*, 53, 2850–2876. <https://doi.org/10.1002/2016WR019185>
- Zha, Y., Yeh, T.-C. J., Illman, W. A., Tanaka, T., Bruines, P., Onoe, H., et al. (2015). What does hydraulic tomography tell us about fractured geological media? A field study and synthetic experiments. *Journal of Hydrology*, 531(1), 17–30. <https://doi.org/10.1016/j.jhydrol.2015.06.013>
- Zha, Y., Yeh, T.-C. J., Illman, W. A., Tanaka, T., Bruines, P., Onoe, H., et al. (2016). An application of hydraulic tomography to a large-scale fractured granite site, Mizunami, Japan. *Ground Water*, 54(6), 793–804. <https://doi.org/10.1111/gwat.12421>
- Zhang, D., & Lu, Z. (2004). An efficient, high-order perturbation approach for flow in random porous media via Karhunen-Loève and polynomial expansions. *Journal of Computational Physics*, 194(2), 773–794. <https://doi.org/10.1016/j.jcp.2003.09.015>
- Zhao, Z., Illman, W. A., Yeh, T.-C. J., Berg, S. J., & Mao, D. (2015). Validation of hydraulic tomography in an unconfined aquifer: A controlled sandbox study. *Water Resources Research*, 51, 4137–4155. <https://doi.org/10.1002/2015WR016910>
- Zhou, Y., & Cardiff, M. (2017). Oscillatory hydraulic testing as a strategy for NAPL source zone monitoring: Laboratory experiments. *Journal of Contaminant Hydrology*, 200, 24–34.
- Zhou, Y., Lim, D., Cupola, F., & Cardiff, M. (2016). Aquifer imaging with pressure waves—Evaluation of low-impact characterization through sandbox experiments. *Water Resources Research*, 52, 2141–2156. <https://doi.org/10.1002/2015WR017751>
- Zhu, J., & Yeh, T.-C. J. (2005). Characterization of aquifer heterogeneity using transient hydraulic tomography. *Water Resources Research*, 41, W07028. <https://doi.org/10.1029/2004WR003790>