

**1 Large-Scale Hydraulic Tomography and Joint  
2 Inversion of Head and Tracer Data using the  
3 Principal Component Geostatistical Approach  
4 (PCGA)**

J. Lee and P. K. Kitanidis<sup>1</sup>

---

J. Lee, Department of Civil and Environmental Engineering, Stanford University, Stanford, 94305, USA. (jonghyun@stanford.edu)

P. K. Kitanidis, Department of Civil and Environmental Engineering, Stanford University, Stanford, 94305, USA. (peterk@stanford.edu)

<sup>1</sup>Department of Civil and Environmental Engineering, Stanford University, Stanford, 94305, USA.

5 **Abstract.** The stochastic geostatistical inversion approach is widely used  
6 in subsurface inverse problems to estimate unknown parameter fields and cor-  
7 responding uncertainty from noisy observations. However, the approach re-  
8 quires a large number of forward model runs to determine the Jacobian or  
9 sensitivity matrix, thus the computational and storage costs become prohibitive  
10 when the number of unknowns,  $m$ , and the number of observations,  $n$  increase.  
11 To overcome this challenge in large-scale geostatistical inversion, the Prin-  
12 cipal Component Geostatistical Approach (PCGA) has recently been devel-  
13 oped as a “matrix-free” geostatistical inversion strategy that avoids the di-  
14 rect evaluation of the Jacobian matrix through the principal components (low-  
15 rank approximation) of the prior covariance and the drift matrix with a finite-  
16 difference approximation. As a result, the proposed method requires about  
17  $K$  runs of the forward problem in each iteration independently of  $m$  and  $n$ ,  
18 where  $K$  is the number of principal components and can be much less than  
19  $m$  and  $n$  for large-scale inverse problems. Furthermore, the PCGA is easily  
20 adaptable to different forward simulation models and various data types for  
21 which the adjoint-state method may not be implemented suitably. In this  
22 paper, we apply the PCGA to representative subsurface inverse problems to  
23 illustrate its efficiency and scalability. The low-rank approximation of the  
24 large-dimensional dense prior covariance matrix is computed through a ran-  
25 domized eigen-decomposition. A hydraulic tomography problem in which the  
26 number of observations is typically large is investigated first to validate the  
27 accuracy of the PCGA compared with the conventional geostatistical approach.

28 Then the method is applied to a large-scale hydraulic tomography with 3 mil-  
29 lion unknowns and it is shown that underlying subsurface structures are char-  
30 acterized successfully through an inversion that involves an affordable num-  
31 ber of forward simulation runs. Lastly, we present a joint inversion of head  
32 and tracer test data using MODFLOW and MT3DMS as coupled black-box  
33 forward simulation solvers. These applications demonstrate the advantages  
34 of the PCGA, *i.e.*, the scalability to high-dimensional inverse problems and  
35 the ability to utilize multiple forward models as black boxes.

## 1. Introduction

36 In the subsurface inverse problem, spatially distributed parameters of interest such  
37 as hydraulic conductivity are characterized from a limited number of hydrogeophysical  
38 measurements like hydraulic heads, contaminant concentration history, or electrical resis-  
39 tivity data. Such problems are ill-posed, *i.e.*, the solution is non-unique or too sensitive  
40 to data, and subject to uncertainty, thus they are commonly evaluated within a statistical  
41 framework [McLaughlin and Townley, 1996; Carrera et al., 2005; Oliver et al., 2008; Kai-  
42 pio and Somersalo, 2007]. Among many inversion approaches, the Bayesian quasi-linear  
43 geostatistical approach [Kitanidis, 1995, 2010] is a versatile method to estimate unknown  
44 subsurface parameters and quantify the corresponding uncertainty rigorously. The geosta-  
45 tistical approach has been applied to many engineering applications such as contaminant  
46 source identification [Snodgrass and Kitanidis, 1997], historical groundwater contaminant  
47 distribution estimation [Michalak and Kitanidis, 2004], tracer data inversion [Nowak and  
48 Cirpka, 2006; Fienen et al., 2009], hydraulic tomography [Li et al., 2007, 2008; Cardiff  
49 et al., 2009; Cardiff and Barrash, 2011], upstream flow estimation in rivers [D’Oria and  
50 Tanda, 2012], atmospheric modeling [Yadav et al., 2010; Miller et al., 2013] and others.

51 With rapid advances in the computing power and sensor technology, large-scale inverse  
52 modeling using large volumes of various types of data, requiring multi-physics model-  
53 ing, has been an active area of research [Flath et al., 2011; Bui-Thanh et al., 2012; Yeh  
54 et al., 2008; Liu and Kitanidis, 2011]. When the geostatistical approach is applied to  
55 high-dimensional problems, two main challenges arise: 1) as the number of unknowns  
56  $m$  increases, the computation and storage costs regarding the dense covariance matrix

57 increase nonlinearly and become prohibitive, 2) when the number of measurements  $n$  in-  
58 creases, the Jacobian (sensitivity) matrix needs  $n$  forward model evaluations using the  
59 adjoint state method, which would be often the most expensive part during the inver-  
60 sion. Building adjoint-state models when multiple processes are involved, such as flow,  
61 reactive transport, electrical resistance, etc., can be quite challenging. Previous efforts  
62 have addressed the former, *i.e.*, the fast computation of dense covariance matrix products,  
63 and fast linear algebra methods such as the fast Fourier transform (FFT)-based approach  
64 [Nowak et al., 2003] and the hierarchical matrix approach [Saibaba and Kitanidis, 2012;  
65 Ambikasaran et al., 2013] reduce the computational and storage costs involving the co-  
66 variance matrix products. However, these approaches still need to access the Jacobian  
67 matrix whose computational cost is prohibitive when one has to employ expensive forward  
68 simulation models and process many measurements during the inversion.

69 In an effort to reduce the computation time associated with forward model runs, Liu  
70 et al. [2013] presented the geostatistical reduced-order inversion (GROM) approach in  
71 which the  $m$ -dimensional unknown parameter space is reparameterized by the solution  
72 space of the geostatistical approach [Kitanidis, 1998], *i.e.*, the  $n + p$  dimensional space  
73 spanned by the  $n$ -column space of the cross-covariance between measurements and un-  
74 knowns and the  $p$ -column space of the drift. In this way, only  $n + p$  full numerical model  
75 simulations are executed at the beginning to construct a reduced-order model and inex-  
76 pensive low-dimensional forward models are used during the inversion. While  $n + p$  full  
77 numerical simulations need to be simulated repeatedly at each iteration when the data  
78 misfit does not decrease, it is observed that the full-model simulations at the beginning are

79 sufficient to obtain acceptable estimates for their laboratory-scale hydraulic tomography  
80 experiments where the measurement error is greater than the approximation error.

81 Kitanidis and Lee [2014, *in review*] presented the theoretical framework of the Principal  
82 Component Geostatistical Approach (PCGA) that avoids the direct evaluation of the  
83 Jacobian matrix through the principal components associated with the prior covariance  
84 and the drift matrix and the finite-difference derivative computation. The method is best  
85 suited for large-scale inverse problems since only a relatively small number of forward  
86 model runs is needed. In addition, the method can be easily adapted to any numerical  
87 simulation models, which is especially beneficial for joint inverse problems that require  
88 coupled numerical simulations.

89 In this paper, we test the efficiency and scalability of the PCGA to challenging subsur-  
90 face inverse problems such as large-scale hydraulic tomography and joint data inversion  
91 of head and tracer test data and provide implementation guidance. In addition, while  
92 our previous work showed the use of a prior ensemble as a driver to decompose the prior  
93 covariance, we here use a different randomized eigen-decomposition method [Halko et al.,  
94 2011] to directly decompose the prior covariance matrix with controlled accuracy. The  
95 randomized technique implements a linear combination of a randomly selected column  
96 space of a matrix to find a near-optimal low-rank approximation of the matrix in a very  
97 efficient manner and has become a popular tool due to its accuracy and easy implemen-  
98 tation over the last decade. With the efficient matrix decomposition tool supported and  
99 maintained by the applied mathematics research community, the PCGA can greatly re-  
100 duce the computational costs without losing the accuracy compared to the conventional  
101 geostatistical approach as shown later.

102 The remainder of this paper is organized as follows. In section 2, we review the quasi-  
103 linear geostatistical approach and the PCGA. In section 3, we present a steady-state  
104 hydraulic tomography example to examine the accuracy and efficiency of our approach.  
105 The method is then applied to a large-scale hydraulic tomography with  $m = 3 \times 10^6$   
106 unknowns to test the scalability of the method. Lastly, the PCGA is applied to a joint  
107 inversion application using hydraulic head and tracer data to investigate the applicability  
108 of our method to general joint inverse problems. Coupled flow and transport simulations  
109 are carried out using the widely-used software tools, MODFLOW [Harbaugh et al., 2000]  
110 and MT3DMS [Zheng and Wang, 1999].

## 2. Overview of the Principal Component Geostatistical Approach

111 In this section, we review the quasi-linear geostatistical approach [Kitanidis, 1995] and  
112 the PCGA [Kitanidis and Lee, 2014]. The observation equation, which relates the  $m \times 1$   
113 vector of unknowns  $\mathbf{s}$  to the  $n \times 1$  vector of the data  $\mathbf{y}$  is

$$114 \quad \mathbf{y} = h(\mathbf{s}) + \mathbf{v} \quad (1)$$

115 where  $h$  is the forward model mapping the parameter space  $\mathbb{R}^m$  to the measurement space  
116  $\mathbb{R}^n$ ,  $\mathbf{v}$  is Gaussian with zero mean and covariance  $\mathbf{R}$  that accounts for errors in the data  
117  $\mathbf{y}$  and the forward model  $h$ . The prior probability of  $\mathbf{s}$  is Gaussian with mean  $\mathbf{X}\boldsymbol{\beta}$ , where  
118  $\mathbf{X}$  is the  $m \times p$  known (polynomial) matrix and  $\boldsymbol{\beta}$  is the  $p \times 1$  unknown vector (typically  
119  $p = 1$ ), and  $\mathbf{Q}$  is the generalized covariance matrix.

120 The posterior pdf of  $\mathbf{s}$  and  $\boldsymbol{\beta}$  are obtained through Bayes theorem and its negative  
121 loglikelihood,  $-\ln p''(\mathbf{s}, \boldsymbol{\beta})$ , is

$$122 \quad -\ln p''(\mathbf{s}, \boldsymbol{\beta}) = \frac{1}{2}(\mathbf{y} - h(\mathbf{s}))^\top \mathbf{R}^{-1}(\mathbf{y} - h(\mathbf{s})) + \frac{1}{2}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{Q}^{-1}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta}) \quad (2)$$

123 By minimizing (2) with respect to  $\mathbf{s}$  and  $\boldsymbol{\beta}$ , we can obtain the maximum a posterior (MAP)  
 124 or most likely value  $\hat{\mathbf{s}}$ , commonly computed through an iterative Gaussian-Newton method  
 125 described as follows.

126 Starting with the latest “best” estimate  $\bar{\mathbf{s}}$ , we update to a new solution until  $\bar{\mathbf{s}}$  converges  
 127 to  $\hat{\mathbf{s}}$ . First, the  $n \times m$  Jacobian or sensitivity matrix  $\mathbf{H}$  of  $h$  at  $\bar{\mathbf{s}}$  is evaluated as:

$$128 \quad \mathbf{H} = \left. \frac{\partial h}{\partial \mathbf{s}} \right|_{\mathbf{s}=\bar{\mathbf{s}}} \quad (3)$$

129 Then, assuming that the actual best estimate  $\hat{\mathbf{s}}$  is close to  $\bar{\mathbf{s}}$ , linearize  $h(\hat{\mathbf{s}})$  as

$$130 \quad h(\hat{\mathbf{s}}) = h(\bar{\mathbf{s}}) + \mathbf{H}(\hat{\mathbf{s}} - \bar{\mathbf{s}}) \quad (4)$$

131 and based on the linearization, the updated solution for the next iteration is computed as

$$132 \quad \bar{\mathbf{s}} = \mathbf{X}\bar{\boldsymbol{\beta}} + \mathbf{Q}\mathbf{H}^\top \bar{\boldsymbol{\xi}} \quad (5)$$

133 where  $\bar{\boldsymbol{\beta}}$  and  $\bar{\boldsymbol{\xi}}$  are computed by solving a single linear system of  $n + p$  equations:

$$134 \quad \begin{bmatrix} \mathbf{H}\mathbf{Q}\mathbf{H}^\top + \mathbf{R} & \mathbf{H}\mathbf{X} \\ (\mathbf{H}\mathbf{X})^\top & \mathbf{0} \end{bmatrix} \begin{bmatrix} \bar{\boldsymbol{\xi}} \\ \bar{\boldsymbol{\beta}} \end{bmatrix} = \begin{bmatrix} \mathbf{y} - h(\bar{\mathbf{s}}) + \mathbf{H}\bar{\mathbf{s}} \\ \mathbf{0} \end{bmatrix} \quad (6)$$

135 The equations (3) - (6) are repeated until  $\bar{\mathbf{s}}$  converges. In most cases, the system (6) can  
 136 be solved using direct solvers such as UMFPACK [Davis, 2004] up to the dimension of  
 137  $n \sim \mathcal{O}(10,000)$  efficiently on current hardware (with  $\mathcal{O}(1)$  GB storage). Larger systems  
 138 can be computed in a matrix-free fashion using iterative Krylov space solvers as described  
 139 in Saibaba and Kitanidis [2012].

140 Note also that the objective function to be minimized can be written as

$$141 \quad J = \frac{1}{2}(\mathbf{y} - h(\mathbf{s}))^\top \mathbf{R}^{-1}(\mathbf{y} - h(\mathbf{s})) + \frac{1}{2}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta})^\top \mathbf{Q}^{-1}(\mathbf{s} - \mathbf{X}\boldsymbol{\beta}) \\ 142 \quad = \frac{1}{2}(\mathbf{y} - h(\mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\mathbf{H}^\top \boldsymbol{\xi}))^\top \mathbf{R}^{-1}(\mathbf{y} - h(\mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\mathbf{H}^\top \boldsymbol{\xi})) + \frac{1}{2}\boldsymbol{\xi}^\top \mathbf{H}\mathbf{Q}\mathbf{H}^\top \boldsymbol{\xi} \quad (7)$$

143 The equation (7) can be used to gauge the progress of minimization and to make sure that  
 144 the new solution is not worse than the previous one. For strongly nonlinear problems,  
 145 the Levenberg-Marquardt method [Nowak and Cirpka, 2004] or a line search [Zanini and  
 146 Kitanidis, 2009; Liu and Kitanidis, 2011] can be adopted for the better convergence.

147 Once the best estimate is determined, uncertainty of the estimate is characterized by  
 148 the posterior covariance matrix  $\mathbf{V}$  computed as:

$$149 \quad \mathbf{V} = \mathbf{Q} - \mathbf{X}\mathbf{M} - \mathbf{Q}\mathbf{H}^\top \mathbf{\Lambda}^\top \quad (8)$$

150 where  $\mathbf{X}$  and  $\mathbf{M}$  are computed from

$$151 \quad \begin{bmatrix} \mathbf{H}\mathbf{Q}\mathbf{H}^\top + \mathbf{R} & \mathbf{H}\mathbf{X} \\ (\mathbf{H}\mathbf{X})^\top & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}^\top \\ \mathbf{M} \end{bmatrix} = \begin{bmatrix} \mathbf{H}\mathbf{Q} \\ \mathbf{X}^\top \end{bmatrix} \quad (9)$$

152 Alternatively, the estimation uncertainty can be quantified by the posterior ensemble,  
 153 *i.e.*, conditional realizations sampled from the posterior pdf using either the direct de-  
 154 composition of the posterior covariance or the parametric bootstrapping sampling method  
 155 [Kitanidis, 1995; Kitanidis and Lee, 2014].

156 The geostatistical approach described above can be applied to small-to-moderate-scale  
 157 inverse problems (up to  $m \sim \mathcal{O}(10^4)$ ). However, when the method is implemented on  
 158 finely resolved grids with a large number of measurements, computational challenges arise  
 159 mainly from the construction of  $\mathbf{H}$  and the matrix products of  $\mathbf{H}$ , *i.e.*,  $\mathbf{H}\mathbf{s}$ ,  $\mathbf{H}\mathbf{X}$ ,  $\mathbf{H}\mathbf{Q}$  and  
 160  $\mathbf{H}\mathbf{Q}\mathbf{H}^\top$ . The Jacobian matrix  $\mathbf{H}$  must be calculated several times and each computation  
 161 using the adjoint-state method, which is recommended when  $n \ll m$ , corresponds to  $n$   
 162 solutions of the forward problem. Usually, this is the most computationally expensive  
 163 part of the geostatistical approach, particularly when many measurements are available  
 164 and several Gaussian-Newton iterations are needed. Additionally, the storage of  $\mathbf{H}$  and

165 its multiplication with  $\mathbf{Q}$  become prohibitive for large  $m$  and  $n$ . Furthermore, the compu-  
 166 tation and storage costs of the posterior covariance matrix may not be tractable in many  
 167 cases.

168 In order to tackle these difficulties, Kitanidis and Lee [2014] propose the PCGA that  
 169 implements the low-rank approximation of  $\mathbf{Q}$  and a finite-difference approximation to  
 170 avoid the direct evaluation of  $\mathbf{H}$ . Assume we approximate  $\mathbf{Q}$  as

$$171 \quad \mathbf{Q} \approx \mathbf{Q}_K = \mathbf{Z}\mathbf{Z}^T = \sum_{i=1}^K \boldsymbol{\zeta}_i \boldsymbol{\zeta}_i^T \quad (10)$$

172 where  $\mathbf{Q}_K$  is a rank- $K$  approximation of  $\mathbf{Q}$ ,  $\mathbf{Z}$  is a  $m \times K$  matrix and  $\boldsymbol{\zeta}_i$  is the  $i$ -th column  
 173 vector of  $\mathbf{Z}$ . Furthermore, a generic Jacobian product  $\mathbf{H}\mathbf{u}$  needed in (4) (*e.g.*,  $\mathbf{u} = \bar{\mathbf{s}}$  or  
 174  $\boldsymbol{\zeta}_i$ ) can be computed approximately at the cost of two forward model evaluations from  
 175 the Taylor series expansion:

$$176 \quad h(\bar{\mathbf{s}} + \delta\mathbf{u}) = h(\bar{\mathbf{s}}) + \delta\mathbf{H}\mathbf{u} + \mathcal{O}(\delta^2) \quad (11)$$

177 where  $\mathbf{u}$  is a  $m \times 1$  vector and  $\delta$  is the finite-difference interval. For example,  $\mathbf{H}\bar{\mathbf{s}}$  can be  
 178 computed from

$$179 \quad \mathbf{H}\bar{\mathbf{s}} = \frac{1}{\delta} [h(\bar{\mathbf{s}} + \delta\bar{\mathbf{s}}) - h(\bar{\mathbf{s}})] + \mathcal{O}(\delta) \approx \frac{1}{\delta} [h(\bar{\mathbf{s}} + \delta\bar{\mathbf{s}}) - h(\bar{\mathbf{s}})] \quad (12)$$

180 As a result, the computational cost associated with a Jacobian-vector product reduces  
 181 from  $n$  to two forward simulations and the storage cost becomes  $\mathcal{O}(mK)$  from  $\mathcal{O}(mn)$ .

182 Similarly,  $\mathbf{H}\mathbf{X}$  can be computed by

$$183 \quad \mathbf{H}\mathbf{X}_i \approx \frac{1}{\delta} [h(\bar{\mathbf{s}} + \delta\mathbf{X}_i) - h(\bar{\mathbf{s}})] \quad (13)$$

184 where  $\mathbf{X}_i$  is the  $i$ -th column of  $\mathbf{X}$ . Next, the matrix-matrix products  $\mathbf{H}\mathbf{Q}$  and  $\mathbf{H}\mathbf{Q}\mathbf{H}^\top$  are  
 185 computed by

$$186 \quad \mathbf{H}\mathbf{Q} \approx \mathbf{H}\mathbf{Q}_K = \mathbf{H} \sum_{i=1}^K \boldsymbol{\zeta}_i \boldsymbol{\zeta}_i^\top = \sum_{i=1}^K (\mathbf{H}\boldsymbol{\zeta}_i) \boldsymbol{\zeta}_i^\top \approx \sum_{i=1}^K \boldsymbol{\eta}_i \boldsymbol{\zeta}_i^\top \quad (14)$$

$$187 \quad \mathbf{H}\mathbf{Q}\mathbf{H}^\top \approx \mathbf{H}\mathbf{Q}_K\mathbf{H}^\top = \sum_{i=1}^K (\mathbf{H}\boldsymbol{\zeta}_i) (\mathbf{H}\boldsymbol{\zeta}_i)^\top \approx \sum_{i=1}^K \boldsymbol{\eta}_i \boldsymbol{\eta}_i^\top \quad (15)$$

189 where

$$190 \quad \boldsymbol{\eta}_i = \mathbf{H}\boldsymbol{\zeta}_i \approx \frac{1}{\delta} [h(\mathbf{s} + \delta\boldsymbol{\zeta}_i) - h(\mathbf{s})] \quad (16)$$

191 respectively. Replacing the explicit construction and multiplication of  $\mathbf{H}$  in (3) - (6) by  
 192 (12) - (15) requires a total of  $K + p + 2$  forward model runs in each iteration.

193 One can also take advantage of the PCGA to quantify the uncertainty of the estimation.  
 194 The diagonal entries of the posterior covariance matrix  $\mathbf{V}$  in (8) are often presented as  
 195 the estimation variance and can be computed without constructing  $\mathbf{V}$  explicitly as

$$196 \quad \mathbf{V}_{ii} = \mathbf{Q}_{ii} - \begin{bmatrix} \mathbf{H}\mathbf{Q}_i \\ \mathbf{X}_i^\top \end{bmatrix}^\top \begin{bmatrix} \mathbf{H}\mathbf{Q}\mathbf{H}^\top + \mathbf{R} & \mathbf{H}\mathbf{X} \\ (\mathbf{H}\mathbf{X})^\top & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{H}\mathbf{Q}_i \\ \mathbf{X}_i^\top \end{bmatrix} \quad (17)$$

197 where  $\mathbf{V}_{ii}$  is the  $i$ -th diagonal element of  $\mathbf{V}$ ,  $\mathbf{Q}_{ii}$  is the  $i$ -th diagonal entry or the prior  
 198 variance of  $i$ -th parameter,  $\mathbf{H}\mathbf{Q}_i$  is the  $i$ -th column of  $\mathbf{H}\mathbf{Q}$ , and  $\mathbf{X}_i^\top$  is the  $i$ -th column of  
 199  $\mathbf{X}^\top$ . The cost of computing the entire variance map of the estimate after obtaining (12) -  
 200 (15) is  $\mathcal{O}(n^2m)$ . One may store and reuse the inverse of the co-kriging matrix for a small  
 201 value  $n \leq \mathcal{O}(1000)$  to reduce the computation time further.

202 Before investigating the computational savings in detail, we first examine the accuracy of  
 203 the approximation. Since the matrix-vector products are replaced by the finite-difference  
 204 approximation, the choice of  $\delta$  is crucial to the success of the PCGA. Assuming  $h$  is smooth  
 205 and second-order differentiable, the optimal choice of the finite-difference interval  $\delta$  for the  
 206 forward operator  $h$  at  $\bar{\mathbf{s}}$  that minimizes both first-order approximation and finite-precision

207 arithmetic errors [Gill et al., 1983] can be computed as

$$208 \quad \hat{\delta}_{h(\bar{\mathbf{s}})} = 2\sqrt{\frac{\epsilon}{|\Phi|}} \quad (18)$$

209 where  $\epsilon$  is a machine epsilon (bound on the relative error due to rounding) and  $\Phi$  is  
210 a representative value of the second-derivative of  $h$  around  $\bar{\mathbf{s}}$ . If we do not have any  
211 information on the second derivative of the forward model, the rule of thumb for choosing  
212  $\delta$  is  $\hat{\delta}_{h(\bar{\mathbf{s}})} \sim \sqrt{\epsilon}$ , *i.e.*,  $10^{-4}$  for the single precision floating format and  $10^{-8}$  for the double  
213 precision floating format. In addition, the generic vector  $\mathbf{u}$  evaluated in the forward model  
214  $h$  should have the same unit as  $\bar{\mathbf{s}}$ , thus the optimal choice of  $\delta$  can be computed as

$$215 \quad \hat{\delta} = \hat{\delta}_{h(\bar{\mathbf{s}})} \frac{\|\bar{\mathbf{s}}\|}{\|\mathbf{u}\|} \quad (19)$$

216 where  $\|\cdot\|$  is the norm operator. While one can use a single value of  $\delta$  when the forward  
217 model outputs vary within a couple of orders of magnitude, several  $\delta$ 's can be implemented  
218 for the evaluation of different rows in the matrix-vector products if needed. For joint  
219 inverse problems using coupled forward models that predict different types of data, a  
220 different  $\delta$  for each type of data can be chosen. One may want to experiment with  
221 a high-order finite-difference method in order to choose an appropriate  $\delta$  for a specific  
222 application.

223 With the optimal choice of  $\delta$ , the accuracy of the approximation also depends on the  
224 approximation of  $\mathbf{Q}$ . While any suitable approximation method can be implemented in  
225 (10) as shown in Kitanidis and Lee [2014], in this study we choose the truncated eigen-  
226 decomposition of  $\mathbf{Q}$  for the optimal approximation or “compression” in terms of the  
227 spectral norm:

$$228 \quad \mathbf{Q} \approx \mathbf{Q}_K = \mathbf{V}_K \mathbf{\Lambda}_K \mathbf{V}_K^\top = \sum_{i=1}^K \zeta_i \zeta_i^\top \quad (20)$$

$$\zeta_i = \sqrt{\lambda_i} \mathbf{V}_i \quad (21)$$

231 where  $\mathbf{\Lambda}_K$  is the  $K \times K$  diagonal matrix whose  $i$ -th diagonal element is the  $i$ -th eigenvalue  
 232  $\lambda_i$  of  $\mathbf{Q}$  sorted in descending order and  $\mathbf{V}_K$  is the  $m \times K$  matrix whose  $i$ -th column  $\mathbf{V}_i$   
 233 is the eigenvector corresponding  $\lambda_i$ . The error of  $K$ -rank approximation is the  $K + 1$ -th  
 234 eigenvalue of  $\mathbf{Q}$ , *i.e.*,  $\|\mathbf{Q} - \mathbf{Q}_K\| = \lambda_{K+1}$ . Then  $K$  may be chosen based on a user-defined  
 235 accuracy such as  $\|\mathbf{Q} - \mathbf{Q}_K\| \ll 1$  or up to the point that the computation resources allow.

236 In fact, the accuracy of the low-rank covariance matrix approximation, *i.e.*, the number  
 237 of eigen-pairs one should keep, depends on the smoothness of covariance kernel [Frauen-  
 238 felder et al., 2005; Schwab and Todor, 2006]. For covariance matrices whose eigenvalues  
 239 decay fast, our method performs best with a small value for  $K \ll n$ . For non-differentiable  
 240 covariance kernels such as the exponential function with a short correlation length, one  
 241 may have to select a large value of  $K$  to ensure an accurate approximation of  $\mathbf{Q}$ . However,  
 242 we are primarily interested in the approximation of  $\mathbf{QH}^\top$  and  $\mathbf{HQH}^\top$ , not  $\mathbf{Q}$  because the  
 243 solution in the geostatistical approach (5) is constrained in the space spanned by the  $n$   
 244 columns of the cross-covariance  $\mathbf{QH}^\top$  and the  $p$  columns of the drift  $\mathbf{HX}$  [Kitanidis, 1998].  
 245 Because the  $n \times m$  Jacobian (sensitivity) matrix  $\mathbf{H}$  is typically low-rank ( $\ll n$ ) reflecting  
 246 the low information content in the measurements, the approximation errors in  $\mathbf{QH}^\top$  and  
 247  $\mathbf{HQH}^\top$  may not be significant with a small value of  $K$ . It is also noted that the error  
 248 matrix  $\mathbf{R}$ , which represents uncertainty in the forward model and measurements, reduces  
 249 the effects of approximation error further as we observe smooth estimates for large mea-  
 250 surement error. Thus the effect of the error introduced by the low-rank covariance and  
 251 finite-difference approximations on the solution would become negligible with a moder-  
 252 ate number of  $K$  in most practical cases. In this paper, we use the relative error of the

253 low-rank approximation as

$$254 \quad \frac{\|\mathbf{Q} - \mathbf{Q}_K\|}{\|\mathbf{Q}\|} = \frac{\lambda_{K+1}}{\lambda_1} \quad (22)$$

255 to determine the number of eigen-modes  $K$  used in the PCGA.

256 However, the cost of the conventional eigen-decomposition in (20) is  $\mathcal{O}(m^2K)$ , which  
257 can be extremely expensive for large-scale applications. For the efficient decomposition  
258 scalable to large-scale problems, we employ the randomized eigen-decomposition [Halko  
259 et al., 2011], a method to approximate the exact truncated eigen-decomposition with  
260 smaller decomposition costs, summarized in Appendix A. The idea of the randomized  
261 method is to use randomly sampled columns of  $\mathbf{Q}$  to capture the dominant eigen-spaces  
262 of  $\mathbf{Q}$ . Combined with fast matrix linear algebra methods such FFT [Nowak et al., 2003],  
263 the fast-multipole method [Greengard and Rokhlin, 1987; Fong and Darve, 2009] or hier-  
264 archical matrices [Ambikasaran et al., 2013; Saibaba and Kitanidis, 2012], one can achieve  
265 a computational cost of  $O(mK \log m)$ . With a certain choice of parameters (see Appendix  
266 A), the estimated error of the randomized technique with a high probability ( $\sim 1 - 10^{-10}$ )  
267 is given by [Halko et al., 2011; Rokhlin et al., 2009]:

$$268 \quad \|\mathbf{Q} - \tilde{\mathbf{Q}}_K\| \leq Cm^{1/6} \lambda_{K+1} \quad (23)$$

269 where  $\tilde{\mathbf{Q}}_K$  is the  $K$ -rank approximation of  $\mathbf{Q}$  using the randomized eigen-decomposition  
270 and  $C$  is a constant. Note that the variance of the error in (23) is practically small. In  
271 Figure 1, the performance of the randomized eigen-decomposition is examined for two  
272 covariance matrices constructed from exponential  $q(x, x') = \exp\left(-\frac{|x-x'|}{0.3}\right)$  and Gaussian  
273 (double-exponential)  $q(x, x') = \exp\left(-\frac{|x-x'|^2}{0.3^2}\right)$  kernels on a 100 by 100 grid in  $x \in [0, 1]^2$ .  
274 50 randomized eigen-decompositions are tested and the results shows clearly that the  
275 approximation error is almost negligible and the accuracy of the randomized method is

276 comparable to the conventional decomposition method. For most problems, the decom-  
277 position is very fast and fits in the RAM memory of contemporary standard PCs. For  
278 example, for a covariance matrix with  $m = 10^6$ , it took about 3 minutes and required 800  
279 MB memory to finish a  $K = 100$ -rank approximation in a laptop equipped with quad-core  
280 2.1 GHz CPU and 8 GB RAM. Thus, one may start the randomized eigen-decomposition  
281 with large  $K$ , e.g.,  $K \sim \mathcal{O}(100)$ , then choose a value for  $K$  that satisfies a reasonable  
282 approximation accuracy.

283 Once we have the decomposition of the prior matrix  $\mathbf{Q}$  in (10), only  $K + p + 2$  forward  
284 runs are needed in each iteration, which can be a significant reduction of computational  
285 costs compared to  $n + 1$  forward model runs with the adjoint state method. Additionally,  
286 the computational costs for matrix-multiplication reduce to  $\mathcal{O}(mK \log m)$  from  $\mathcal{O}(m^2n)$  in  
287 the conventional geostatistical approach or  $\mathcal{O}(mn \log m)$  in the approach with fast linear  
288 algebra methods. A comparison of the computational costs is presented in Table 1. One  
289 could achieve further computation savings by starting with a small value  $K$  during the  
290 first few iterations or until the minimization of the objective value in (7) does not progress  
291 then increasing  $K$  later because inaccurate approximation may be sufficient to compute  
292 Gauss-Newton steps when the initial guess is far from the minimum.

293 Besides the improvement in the computation and storage costs, additional advantages  
294 can be expected from the PCGA due to its simplicity and easy of implementation. In  
295 general, the development and implementation of the adjoint state method for coupled  
296 nonlinear and/or time-dependent forward models is challenging and time-consuming be-  
297 cause a formulation of adjoint equations and a differentiation of each nonlinear forward  
298 operator are required and a massive amount of transient solutions may have to be stored

299 for a time-dependent adjoint calculation. Rewriting the adjoint-state codes from existing  
300 complex forward model code would also take a large amount of time and effort. In some  
301 cases, one might prefer efficient, stable and highly optimized forward solvers that do not  
302 allow to access the code. The PCGA can deal with such difficulties effectively.

### 3. Numerical Experiments

303 We demonstrate the performance of the PCGA on two applications - a steady-state hy-  
304 draulic tomography using sequential pumping tests and a joint inversion using hydraulic  
305 head and tracer test data. In the first application, we start with a moderate number of un-  
306 knowns and compare the results with those obtained from the conventional geostatistical  
307 approach. Then a large-scale hydraulic tomography problem with three million unknowns  
308 is investigated. In the second application, a joint inversion problem with hydraulic head  
309 and tracer test data is examined. The PCGA and a 2-D steady-state groundwater flow  
310 model used in the first two applications were implemented in MATLAB. The coupled flow  
311 and transport for the joint inversion were simulated using MODFLOW [Harbaugh et al.,  
312 2000] and MT3DMS [Zheng and Wang, 1999]. The coupling was easily done using mflab,  
313 an efficient MATLAB interface (<http://code.google.com/p/mflab/>). All test cases were  
314 run on a modest workstation equipped with 12-core 3 GHz processors and 128 GB RAM.

#### 3.1. Application 1 : Hydraulic Tomography

315 In this section, we consider a steady-state hydraulic tomography problem to examine  
316 the accuracy of our method compared with the textbook approach. By the textbook  
317 approach, we here mean the conventional geostatistical approach with the adjoint state  
318 method and the fast-linear algebra implementation. A synthetic two-dimensional test case

319 adapted from Zanini and Kitanidis [2009] is used as shown in Figure 2 and all parameters  
320 used in this application are listed in Table 2. The governing equation for the steady-state  
321 depth-integrated flow in a confined aquifer is

$$322 \quad -\nabla \cdot (T\nabla h) = N + Q_i\delta(x - x_i) \quad (24)$$

323 with the following boundary conditions:

$$324 \quad h = h_D, \text{ on } \Gamma_D \quad (25)$$

$$325 \quad n \cdot T\nabla h = 0, \text{ on } \Gamma_N \quad (26)$$

327 where  $T$  is the transmissivity,  $h$  is the hydraulic head,  $N$  is the recharge rate,  $q_i$  is the  
328 pumping rate for a pumping test at a well location  $x_i$  marked with a circle in Figure  
329 2,  $\delta(x)$  is the Dirac delta function and  $n$  is the unit vector normal to the boundaries.  
330 A constant-head boundary condition on the boundaries of  $x = 0$  and 1000 m ( $\Gamma_D$ ) and  
331 no-flux condition on  $y = 0$  and 750 m ( $\Gamma_N$ ) are considered. Two “True” log-transmissivity  
332 ( $\ln T$ ) fields shown in Figure 3 were generated on the  $100 \times 75$  grid using isotropic cubic  
333 generalized covariance and exponential covariance models representing a high-contrast  
334 smooth field (Case 1) and a highly heterogeneous field (Case 2), respectively. Based on  
335 the “true”  $\ln T$  fields, sequential pumping tests were carried out numerically at each well  
336 location and hydraulic heads were measured at the well locations except the pumping well  
337 resulting a total of 870 head measurements from 30 pumping tests with 29 observations  
338 per test. Therefore, we estimate 7,500  $\ln T$  values from 870 head observations. A Gaussian  
339 error with zero mean and the standard deviation of 0.5 m was added to the simulated  
340 measurements and this error corresponds to 5% of the maximum drawdown in the domain.  
341 The finite-difference interval  $\delta$  was chosen from (19) as  $\hat{\delta} = 10^{-6}$  based on the double

342 precision machine epsilon  $\epsilon \approx 10^{-16}$  and  $h'' \approx N/T$ . We employ isotropic exponential  
343 covariance models to estimate  $\ln T$ , and the structural parameters for variogram and  
344 measurement error were determined using the cR/Q2 criteria [Kitanidis, 1997]. The initial  
345 guess was set to a homogeneous field of  $\ln T = 2$  for all the considered cases. The solutions  
346 in all tests converged within 10 iterations.

347 The best estimates of the log-transmissivity distribution for  $K = 25, 50$ , and  $75$  are  
348 shown in Figure 4. For the comparison purpose, the textbook approach was applied to  
349 both cases in order to use the estimate as the “reference” (Figure 4 (d) and (h)). In the first  
350 case, the reference field in Figure 4 (d) captures most features of the true  $\ln T$  field since the  
351 measurements are enough to estimate the smooth field generated from the differentiable  
352 cubic covariance model. As expected, the PCGA performs well, even the estimate with  
353  $K = 25$  in Figure 4 (a) is very similar to the reference estimate. The spectrum of the  
354 prior matrix  $\mathbf{Q}$  shown in Figure 5 (black) indicates that the approximation error becomes  
355 negligible with a moderate number of  $K > 50$ . In the second case, on the other hand,  
356 the best estimate in Figure 4 (h) identifies only large-scale features of the true  $\ln T$  field  
357 in Figure 3 because the measurements cannot resolve the small-scale details of the highly  
358 heterogenous true field. For the PCGA,  $K = 50$  is required to obtain a close estimate  
359 (Figure 4 (f)) to the reference. The eigenvalues of the prior covariance  $\mathbf{Q}$  decay slowly as  
360 in Figure 5 (red) so that we need to keep more eigen-pairs to reduce the approximation  
361 error. Still, the number of forward runs in the PCGA is dramatically reduced compared  
362 to the textbook approach that requires  $n = 871$  forward runs in each iteration to evaluate  
363 the full Jacobian  $\mathbf{H}$ . These results indicate that even for strongly heterogeneous  $\ln T$  field,  
364 we can achieve significant computational savings.

365 Figure 6 shows the corresponding variance of the  $\ln T$  estimation. The uncertainty  
 366 nearby the observation well is found to be low, while high uncertainty is observed in the  
 367 region where large drawdowns take place. Similar to the best estimates, the variance  
 368 estimates using  $K = 50$  or larger are close to the reference estimates for both cases.

369 In Figure 7 (a) and (b), we investigate the effect of  $K$  to the accuracy of estimates in  
 370 terms of element-wise root mean square error (RMSE) defined as

$$371 \quad RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m \left( \ln T_i^{ref} - \ln T_i^{est} \right)^2} \quad (27)$$

372 where  $\ln T_i^{ref}$  is the log-transmissivity of the reference estimate in the grid cell  $i$ , and  $\ln T_i^{est}$   
 373 is the  $i$ -th log-transmissivity estimated by the PCGA in the grid  $i$ . The accuracy of the  
 374 estimation variance is also investigated in Figure 7 (c) and (d). Each test was executed five  
 375 times and we confirmed that the variability from randomized eigen-decomposition does  
 376 not affect the accuracy of the final estimates. For both cases,  $K \geq 100$  gives negligible  
 377 errors ( $RMSE \leq 0.01$ ) to the reference estimate. Since the approximation error comes  
 378 mostly from the covariance approximation, the spectrum analysis shown in Figure 5 can  
 379 be implemented as a priori performance indicator to find an appropriate  $K$ . For this  
 380 application, we observe that a low-rank approximation with about 1% relative error in  $\mathbf{Q}$   
 381 is sufficient to obtain acceptable results ( $RMSE < 0.1$ ), *e.g.*,  $K = 31$  for Case 1 and  $K$   
 382  $= 49$  for Case 2.

### 3.2. Application 2 : Large-scale hydraulic tomography with $m = 3 \times 10^6$

383 To investigate the scalability of the PCGA, the number of  $\ln T$  values is set to  $3 \times 10^6$   
 384 by finer discretization of the previous domain, while other parameters are kept the same.  
 385 The “true” transmissivity fields used in the previous section were downscaled by a spline

386 interpolation correspondingly creating the  $2000 \times 1500$  grids as shown in Figure 8 (a) and  
387 (b). 870 noisy measurements were generated from the true  $\ln T$  fields. The exponential  
388 covariance model with the same structural parameters found in the previous section was  
389 used in the inversion. The leading 50 eigenvalues for the two cases shown in Figure 9 were  
390 computed to find the number of eigenmodes  $K$  used in this study. Based on the spectrum  
391 of eigenvalues,  $K = 36$  for the first case and  $K = 48$  for the second case were chosen in  
392 order to take advantage of parallelization with 12 processors. The computation time for  
393 the eigen-decomposition of the  $3 \times 10^6$  by  $3 \times 10^6$  posterior covariance was 3 minutes and  
394 1.5 GB RAM of the storage for  $3 \times 10^6 \times 50$  elements was required.

395 The best estimates for the two cases and their corresponding variance maps are shown  
396 in Figure 8 (c) - (f). The estimates for both cases are practically similar to the true  $\ln T$   
397 fields as observed in the previous case. In Figure 10, the fitting between simulated and  
398 measured hydraulic heads (blue) is plotted for both cases. For the comparison purpose,  
399 hydraulic heads (black) simulated from the mean log-transmissivity field  $\ln(\bar{T}) = 2.5$  is  
400 also plotted to show the quality of the inversion results. Even with a relative small number  
401 of forward model runs, our approach is still able to estimate  $\ln T$  and fit measurement  
402 data very well. Note that because of the downscaling by the interpolation, the details in  
403 the true field was slightly changed compared to the previous case in Figure 3, especially  
404 for the second case, resulting in a larger variance of log-transmissivity. The estimation  
405 results would be improved further if one uses different structural parameters based on the  
406 cR/Q2 criteria.

407 To reduce the computation time, we start with a small value of  $K$  ( $K_{init} = 12$ ), and  $K$   
408 is increased to the chosen value after the first two iterations. By doing so, we could save a

409 significant amount of computational costs while the resulting images were similar to those  
410 obtained using constant  $K$  throughout the iterations. The best estimates were obtained in  
411 7 iterations, spending about 43 minutes with a total of 256 forward runs for the first case  
412 and about 71 minutes with 328 forward runs for the second case. The variance of the best  
413 estimate for each case was computed in about 5 minutes. The results demonstrate that  
414 the PCGA scales well with large-scale problems using an affordable number of forward  
415 model simulations independent of the problem size and the number of measurements.

### 3.3. Application 3 : Joint inversion with head and tracer travel-time data using black-box forward models

416 In this section, we examine the applicability of the PCGA to general inverse problems  
417 that incorporate different types of measurements and require complicated coupled sim-  
418 ulation models. A typical example of this type is the joint inversion of hydraulic head  
419 and tracer test data [Harvey and Gorelick, 1995; Cirpka and Kitanidis, 2000; Nowak and  
420 Cirpka, 2006], which uses coupled flow and transport simulations to incorporate hydraulic  
421 head and tracer concentration measurements to infer subsurface properties. The goal of  
422 this study is to test our method as a generic tool for solving joint inverse problems by  
423 treating coupled simulation models as a black-box model. The widely used numerical sim-  
424 ulation tools MODFLOW-2000 [Harbaugh et al., 2000] and MT3DMS [Zheng and Wang,  
425 1999] are linked to the PCGA inversion code to obtain head and concentration simulation  
426 data from a  $\ln T$  distribution.

427 We consider an artificial two-dimensional test case adapted from Cirpka and Kitanidis  
428 [2000]. All parameters used in this study are listed in Table 3. In an aquifer with the  
429 dimension of  $L_x = 40$  m and  $L_y = 20$  m, an injection well with the flow rate  $Q_w$  is located

430 at  $x_w = 10$  m,  $y_w = 10$  m. No-flow boundaries are considered at  $y = 0$  m and  $y = 20$  m  
 431 while fixed heads  $h_{x=0} = 0.4$  m and  $h_{x=40} = 0$  m are used at the boundaries at  $x = 0$  m  
 432 and  $x = 40$  m. A “true” log-transmissivity ( $\ln T$ ) field was generated using an anisotropic  
 433 exponential covariance function and is shown in Figure 11 (a). Using the true  $\ln T$  field,  
 434 we simulated steady-state flow using MODFLOW-2000 and transient transport of a tracer  
 435 introduced into the injection well at  $(x_w, y_w)$  using MT3DMS. The steady-state head data  
 436 was collected at  $x = 5, 10, 15, 20, 25, 30, 35$  m and  $y = 4, 8, 10, 12, 16$  m. The tracer of  
 437 concentration  $C = 1$  was injected over a period of an hour at the injection well, and the  
 438 transient concentrations were measured every hour for 10 days at  $x = 10, 15, 20, 25, 30,$   
 439  $35$  m and  $y = 8, 10, 12$  m except at the injection well. No-mass flux boundary condition  
 440 is assumed at  $y = 0$  m and  $y = 20$  m and advective boundary condition is imposed at  
 441  $x = 0$  and  $x = 40$ . Instead of the entire concentration history, the mean travel time  
 442 of the tracer data was used in the inversion because temporal moments of the transient  
 443 tracer inversion data is better suited for the tracer inversion [Harvey and Gorelick, 1995;  
 444 Ezzedine and Rubin, 1996; Cirpka and Kitanidis, 2000]. The mean travel time  $\bar{t}_{x_0}$  at the  
 445 monitoring location  $x_0$  is given by

$$\bar{t}_{x_0} = \frac{\int_0^\infty t C(x_0, t) dt}{\int_0^\infty C(x_0, t) dt} \quad (28)$$

447 where  $C(x, t)$  is the concentration at time  $t$ . (28) was computed numerically from dis-  
 448 crete concentration values at each monitoring location. The random error with standard  
 449 deviation of 0.005 m of the head value and 10 % of the mean travel time was added to  
 450 the head and tracer travel time data, respectively. The domain is discretized into 201 by  
 451 101 grid cells, thus we estimate 20301 (201 x 101)  $\ln T$  values using  $n_h = 35$  head and  $n_t$   
 452  $= 17$  travel time data while other parameters are assumed to be known.

453 Since MODFLOW-2000 stores simulation results in the single precision floating point  
454 format, the finite-difference interval for the head data,  $\delta_h$ , is chosen as  $5 \times 10^{-3}$  using  
455 (19) while the interval for tracer data,  $\delta_C$ , is set to the square root of the machine ep-  
456 silon, *i.e.*,  $5 \times 10^{-4}$ . Setting a different interval for each simulation program requires  $K$   
457 MODFLOW executions to compute Jacobian products with respect to head data and  
458 another  $K$  MODFLOW and MT3DMS runs for the travel time. Based on the eigenspec-  
459 tum analysis,  $K = 96$  is chosen that gives the relative approximation error of 2%. The  
460 structural parameters of the covariance matrix and the measurement errors were identical  
461 to those used to create the true solution. The initial guess  $\ln T = -7$  was used and the  
462 estimates converged in 5 iterations for the head data inversion and 7 iterations for the  
463 joint inversion.

464 The best estimate of the log-transmissivity field using the head data alone is shown in  
465 Figure 11 (b). Because of sparse and noisy head observations, the best estimate charac-  
466 terizes the large-scale features of the true field while details at a smaller scale cannot be  
467 recovered. The corresponding variance of estimation in Figure 12 indicates low estimation  
468 variance around the measurement locations while high uncertainty near the boundaries is  
469 observed.

470 In Figure 11, the best estimate using both head and mean travel time data shows a  
471 clear improvement compared to the head data inversion result. Within the plume, the  
472 small-scale features are detected relatively well while the large-scale features outside the  
473 plume are still captured in the estimate. This result is expected because the tracer data is  
474 only informative within the plume due to the advection-dominant process. The variance  
475 of estimation also shows the smaller error bounds in the interior of the plume compared

476 to the variance map for head-data inversion. In Figure 13, we display the fitting between  
477 the simulated and measured data for the two cases. It can be concluded that the use of  
478 travel time data improves the estimation of  $\ln T$ . These results match what is reported in  
479 Cirpka and Kitanidis [2000] and confirm that the PCGA can be applied to joint inversion  
480 problems without difficulty.

481 Note that the geostatistical approach using the temporal moments [Cirpka and Kitani-  
482 dis, 2000] solves steady-state equations to obtain the Jacobian matrix. The inversion  
483 would require  $n_h + 1 = 36$  flow simulations for head data and  $n_h + 1 = 36$  flow simula-  
484 tions plus  $2n_t + 2 = 36$  steady-state transport simulations for first-moment measurements  
485 in each iteration, which needs much smaller computational efforts than  $2K = 198$  flow  
486 simulations and  $K = 96$  transient transport simulations required in our method for this  
487 application. However, the efficiency of the temporal-moment based inversion method can  
488 be degraded in general cases; for instance, if the injection is not uniform along the entire  
489 inflow boundary, an additional transport simulation is required for the zero-th temporal  
490 moment computation. The PCGA is flexible in that it can deal with any practical and  
491 complex conditions, *e.g.*, time-varying boundary conditions, multiple tracer injection or  
492 various tracer injection schemes such as continuous or time-varying tracer injection. When  
493 one has more travel-time measurements from additional monitoring wells or chooses to  
494 add high-order temporal moments, the PCGA still requires the same number of forward  
495 model runs because the number of forward runs is determined independent of the num-  
496 ber of measurements. Furthermore, our proposed method can include additional sources  
497 of information without difficulty, for example, we only need to add a reactive-transport  
498 module in the forward simulation for multiple reactive tracer injection tests.

#### 4. Concluding remark

499 In this work, we have demonstrated the efficiency and accuracy of the PCGA [Kitani-  
500 dis and Lee, 2014 *in review*] on the large-scale and complex joint geostatistical inversion  
501 problems. The randomized eigen-decomposition makes the direct decomposition of the  
502 dense prior covariance matrix feasible and, combined with the finite-difference approxima-  
503 tion, we avoid the computation of the complete Jacobian matrix. The number of forward  
504 model simulations can be determined by users based on the eigen-spectrum of the prior  
505 covariance or their computational resources and budgets (*cf.* [Leube et al., 2013]). The  
506 numerical experiments presented in this paper show that the PCGA can be suitable for  
507 the large-scale problem and be used as a general inversion tool with other black-block  
508 simulation models.

#### Appendix A: Randomized decomposition approach

509 We summarize the randomized technique to compute the eigen-decomposition of the  
510 covariance matrix  $\mathbf{Q}$  [Halko et al., 2011]. First, we would like to find a  $m$  by  $K$  orthonormal  
511 matrix  $\mathbf{W}$  that makes  $\mathbf{W}\mathbf{W}^\top\mathbf{Q}$  close to  $\mathbf{Q}$

$$512 \quad \mathbf{Q} \approx \mathbf{W}\mathbf{W}^\top\mathbf{Q} \quad (\text{A1})$$

513 Assuming we have a good  $\mathbf{W}$ , we can compute the singular value decomposition of the  
514  $K \times m$  matrix  $\mathbf{W}^\top\mathbf{Q}$

$$515 \quad \mathbf{W}^\top\mathbf{Q} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top \quad (\text{A2})$$

516 where  $\mathbf{U}$  is a  $K \times K$  left-singular matrix,  $\mathbf{\Sigma}$  is a  $K \times K$  diagonal matrix whose diagonal  
517 elements are singular value of  $\mathbf{W}^\top\mathbf{Q}$ , and  $\mathbf{V}$  is a  $m \times K$  right-singular matrix. Now we

518 have a low-rank approximation of  $\mathbf{Q}$ :

$$519 \quad \mathbf{Q} \approx \mathbf{W}\mathbf{W}^\top \mathbf{Q} = \mathbf{W}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top \approx \mathbf{V}\mathbf{\Sigma}\mathbf{V}^\top \quad (\text{A3})$$

520 Note that we assume the numerical error is negligible here so that we obtain  $\mathbf{W}\mathbf{U} =$   
521  $\mathbf{V}$ . Please refer to Halko et al. [2011] to construct exact symmetric matrices. In the  
522 randomized approach, we use random (Gaussian) linear combinations of columns of  $\mathbf{Q}$  to  
523 determine  $\mathbf{W}$ :

$$524 \quad \mathbf{W} \subseteq \text{R}(\mathbf{Q}\mathbf{\Omega}) \quad (\text{A4})$$

525 where  $\text{R}$  is the range of the space and  $\mathbf{\Omega}$  is a  $m \times k$  random i.i.d Gaussian matrix, *i.e.*,  
526  $\Omega_{ij} \sim N(0, 1)$ . One can use the QR factorization of  $\mathbf{Q}\mathbf{\Omega}$  to obtain an orthonormal matrix  
527  $\mathbf{W}$ . However, when the eigenvalues of the prior matrix  $\mathbf{Q}$  decay slowly, the accuracy of  
528 the presented algorithm above is degraded significantly. This can be fixed by increasing  
529 the weights on the dominant eigenvalues:

$$530 \quad \mathbf{W} \subseteq \text{R}(\mathbf{Q}^q \mathbf{\Omega}) \quad (\text{A5})$$

531 We also increase the accuracy of the decomposition by finding  $K + p$  eigen-modes, *i.e.*, use  
532 the  $m$  by  $K + p$  orthonormal matrix  $\mathbf{W}$  where  $p$  is the over-sampling parameter ( $p = 15$   
533 is used in this paper). The expected error bound in (23) is computed for  $q = 3$ .

534 **Acknowledgments.** The research was funded by the National Science Foundation  
535 through its ReNUWit Engineering Research Center ([www.renuwit.org](http://www.renuwit.org); NSF EEC-  
536 1028968) and the Charles H. Leavell Graduate Student Fellowship.

## References

- 537 S. Ambikasaran, J. Y. Li, P. K. Kitanidis, and E. Darve. Large-scale stochastic linear  
538 inversion using hierarchical matrices. *Computational Geosciences*, 17(6):913–927, 2013.
- 539 T. Bui-Thanh, C. Burstedde, O. Ghattas, J. Martin, G. Stadler, and L. C. Wilcox.  
540 Extreme-scale UQ for Bayesian inverse problems governed by PDEs. *2012 Interna-  
541 tional Conference for High Performance Computing, Networking, Storage and Analysis  
542 (Sc)*, 2012.
- 543 M. Cardiff and W. Barrash. 3-D transient hydraulic tomography in unconfined aquifers  
544 with fast drainage response. *Water Resources Research*, 47, 2011.
- 545 M. Cardiff, W. Barrash, P. K. Kitanidis, B. Malama, A. Revil, S. Straface, and E. Rizzo.  
546 A potential-based inversion of unconfined steady-state hydraulic tomography. *Ground  
547 Water*, 47(2):259–270, 2009.
- 548 J. Carrera, A. Alcolea, A. Medina, J. Hidalgo, and L. J. Slooten. Inverse problem in  
549 hydrogeology. *Hydrogeology Journal*, 13(1):206–222, 2005.
- 550 O. A. Cirpka and P. K. Kitanidis. Sensitivity of temporal moments calculated by the  
551 adjoint-state method and joint inverting of head and tracer data. *Advances in Water  
552 Resources*, 24(1):89–103, 2000.
- 553 T. A. Davis. Algorithm 832: Umfpack v4.3 - an unsymmetric-pattern multifrontal method.  
554 *ACM Transactions on Mathematical Software*, 30(2):196–199, 2004.
- 555 M. D’Oria and M. G. Tanda. Reverse flow routing in open channels: A Bayesian geosta-  
556 tistical approach. *Journal of Hydrology*, 460:130–135, 2012.
- 557 S. Ezzedine and Y. Rubin. A geostatistical approach to the conditional estimation of  
558 spatially distributed solute concentration and notes on the use of tracer data in the

559 inverse problem. *Water Resources Research*, 32(4):853–861, 1996.

560 M. Fienen, R. Hunt, D. Krabbenhoft, and T. Clemo. Obtaining parsimonious hydraulic  
561 conductivity fields using head and transport observations: A bayesian geostatistical  
562 parameter estimation approach. *Water Resources Research*, 45, 2009.

563 H. P. Flath, L. C. Wilcox, V. Akcelik, J. Hill, B. V. Waanders, and O. Ghattas. Fast  
564 algorithms for bayesian uncertainty quantification in large-scale linear inverse problems  
565 based on low-rank partial hessian approximations. *Siam Journal on Scientific Comput-*  
566 *ing*, 33(1):407–432, 2011.

567 W. Fong and E. Darve. The black-box fast multipole method. *Journal of Computational*  
568 *Physics*, 229(23):8712–8725, 2009.

569 P. Frauenfelder, C. Schwab, and R. A. Todor. Finite elements for elliptic problems with  
570 stochastic coefficients. *Computer Methods in Applied Mechanics and Engineering*, 194  
571 (2-5):205–228, 2005.

572 P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. Computing forward-difference  
573 intervals for numerical optimization. *Siam Journal on Scientific and Statistical Com-*  
574 *puting*, 4(2):310–321, 1983.

575 L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *Journal of*  
576 *Computational Physics*, 73(2):325–348, 1987.

577 N. Halko, P. G. Martinsson, and J. A. Tropp. Finding structure with randomness: Proba-  
578 bilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*,  
579 53(2):217–288, 2011.

580 A. W Harbaugh, E. R Banta, M. C Hill, and M. G McDonald. Modflow-2000, the US  
581 geological survey modular ground-water model: User guide to modularization concepts

582 and the ground-water flow process. Technical Report 2000-92, Geological Survey (U.S.),  
583 2000.

584 C. F. Harvey and S. M. Gorelick. Mapping hydraulic conductivity - sequential conditioning  
585 with measurements of solute arrival time, hydraulic-head, and local conductivity. *Water*  
586 *Resources Research*, 31(7):1615–1626, 1995.

587 J. Kaipio and E. Somersalo. Statistical inverse problems: Discretization, model reduction  
588 and inverse crimes. *Journal of Computational and Applied Mathematics*, 198(2):493–  
589 504, 2007.

590 P. K. Kitanidis. Quasi-linear geostatistical theory for inversing. *Water Resources Research*,  
591 31(10):2411–2419, 1995.

592 P. K. Kitanidis. *Introduction to geostatistics : applications to hydrogeology*. Cambridge  
593 University Press, New York, 1997.

594 P. K. Kitanidis. How observations and structure affect the geostatistical solution to the  
595 steady-state inverse problem. *Ground Water*, 36(5):754–763, 1998.

596 P. K. Kitanidis. Bayesian and geostatistical approaches to inverse problems. In *Large-*  
597 *Scale Inverse Problems and Quantification of Uncertainty*, pages 71–85. John Wiley &  
598 Sons, Ltd, 2010.

599 P. K. Kitanidis and J. Lee. Principal component geostatistical approach for large-  
600 dimensional inverse problems. *Water Resources Research*, 2014. (submitted)

601 P. C. Leube, F. P. J. de Barros, W. Nowak, and R. Rajagopal. Towards optimal allocation  
602 of computer resources: Trade-offs between uncertainty quantification, discretization and  
603 model reduction. *Environmental Modelling & Software*, 50:97–107, 2013.

604 W. Li, A. Englert, O. A. Cirpka, J. Vanderborght, and H. Vereecken. Two-dimensional  
605 characterization of hydraulic heterogeneity by multiple pumping tests. *Water Resources*  
606 *Research*, 43(4), 2007.

607 W. Li, A. Englert, O. A. Cirpka, and H. Vereecken. Three-dimensional geostatistical  
608 inversion of flowmeter and pumping test data. *Ground Water*, 46(2):193–201, 2008.

609 X. Liu and P. K. Kitanidis. Large-scale inverse modeling with an application in hydraulic  
610 tomography. *Water Resources Research*, 47, 2011.

611 X. Liu, Q. Zhou, J. Birkholzer, and W. A. Illman. Geostatistical reduced-order models in  
612 underdetermined inverse problems. *Water Resources Research*, 49(10):6587–6600, 2013.

613 D. McLaughlin and L. R. Townley. A reassessment of the groundwater inverse problem.  
614 *Water Resources Research*, 32(5):1131–1161, 1996.

615 A. M. Michalak and P. K. Kitanidis. Estimation of historical groundwater contaminant  
616 distribution using the adjoint state method applied to geostatistical inverse modeling.  
617 *Water Resources Research*, 40(8), 2004.

618 S. M. Miller, S. C. Wofsy, A. M. Michalak, E. A. Kort, A. E. Andrews, Sebastien C Biraud,  
619 Edward J Dlugokencky, Janusz Eluszkiewicz, Marc L Fischer, and Greet Janssens-  
620 Maenhout. Anthropogenic emissions of methane in the united states. *Proceedings of*  
621 *the National Academy of Sciences*, 110(50):20018–20022, 2013.

622 W. Nowak and O. A. Cirpka. A modified Levenberg-Marquardt algorithm for quasi-linear  
623 geostatistical inversing. *Advances in Water Resources*, 27(7):737–750, 2004.

624 W. Nowak, S. Tenkleve, and O. A. Cirpka. Efficient computation of linearized cross-  
625 covariance and auto-covariance matrices of interdependent quantities. *Mathematical*  
626 *Geology*, 35(1):53–66, 2003.

627 W. Nowak and O. A. Cirpka. Geostatistical inference of hydraulic conductivity and  
628 dispersivities from hydraulic heads and tracer data. *Water Resources Research*, 42  
629 (8):W08416, 2006.

630 D. S. Oliver, A. C. Reynolds, and N. Liu. *Inverse theory for petroleum reservoir charac-*  
631 *terization and history matching*. Cambridge University Press, Cambridge ; New York,  
632 2008.

633 V. Rokhlin, A. Szlam, and M. Tygert. A randomized algorithm for principal component  
634 analysis. *Siam Journal on Matrix Analysis and Applications*, 31(3):1100–1124, 2009.

635 A. K. Saibaba and P. K. Kitanidis. Efficient methods for large-scale linear inversion using  
636 a geostatistical approach. *Water Resources Research*, 48, 2012.

637 C. Schwab and R. A. Todor. Karhunen-loeve approximation of random fields by gen-  
638 eralized fast multipole methods. *Journal of Computational Physics*, 217(1):100–122,  
639 2006.

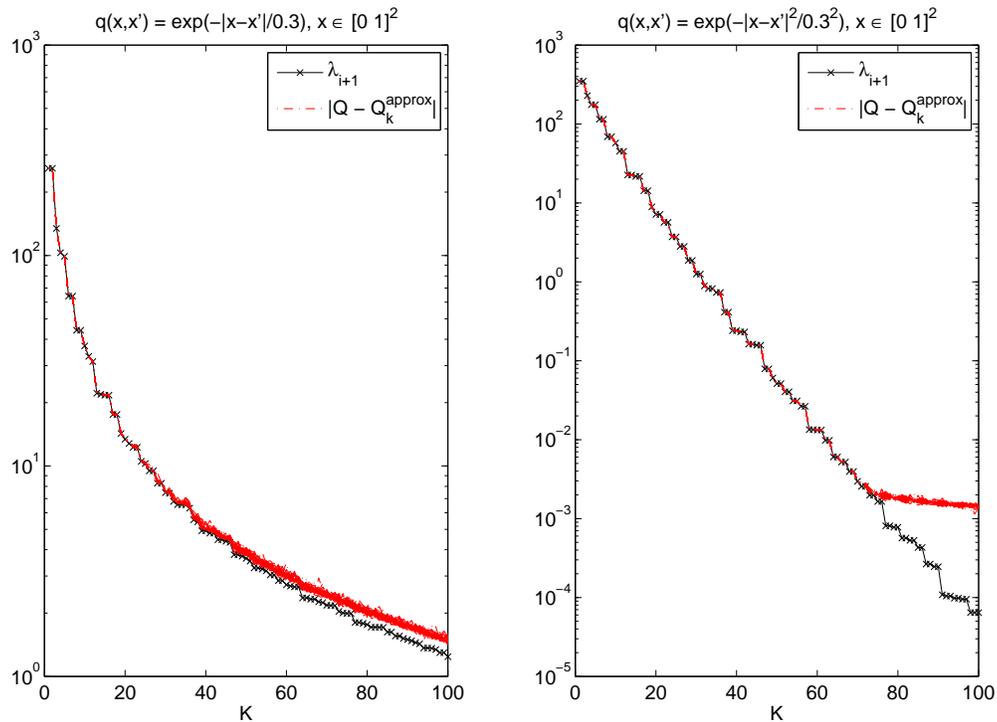
640 M. F. Snodgrass and P. K. Kitanidis. A geostatistical approach to contaminant source  
641 identification. *Water Resources Research*, 33(4):537–546, 1997.

642 V. Yadav, K. L. Mueller, D. Dragoni, and A. M. Michalak. A geostatistical synthesis  
643 study of factors affecting gross primary productivity in various ecosystems of North  
644 America. *Biogeosciences Discussions*, 7(1), 2010.

645 T. C. J. Yeh, C. H. Lee, K. C. Hsu, W. A. Illman, W. Barrash, X. Cai, J. Daniels,  
646 E. Sudicky, L. Wan, G. M. Li, and C. L. Winter. A view toward the future of subsurface  
647 characterization: Cat scanning groundwater basins. *Water Resources Research*, 44(3),  
648 2008.

649 A. Zanini and P. K. Kitanidis. Geostatistical inversing for large-contrast transmissivity  
650 fields. *Stochastic Environmental Research and Risk Assessment*, 23(5):565–577, 2009.

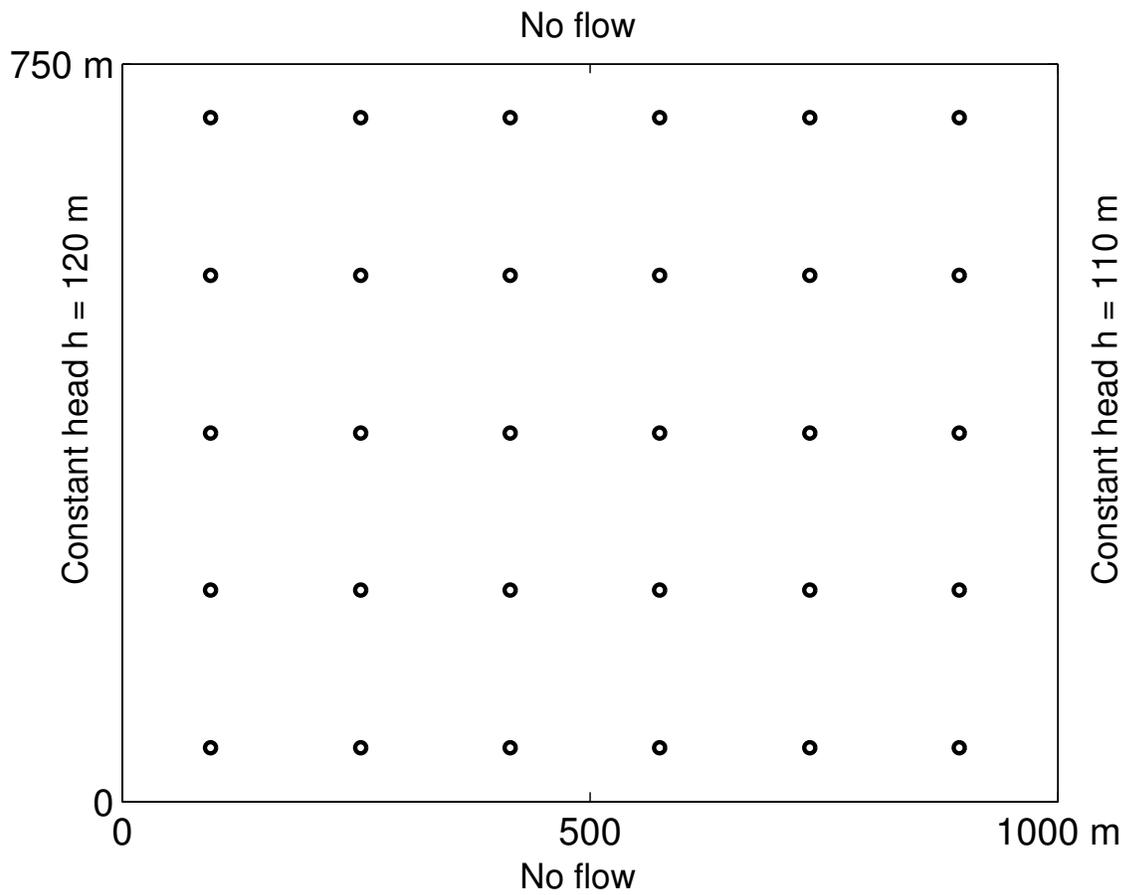
651 Chunmiao Zheng and P Patrick Wang. MT3DMS: A modular three-dimensional multi-  
652 species transport model for simulation of advection, dispersion, and chemical reactions  
653 of contaminants in groundwater systems; documentation and user’s guide. Technical  
654 Report SERDP-99-1, DTIC Document, Vicksburg, Mississippi, 1999.



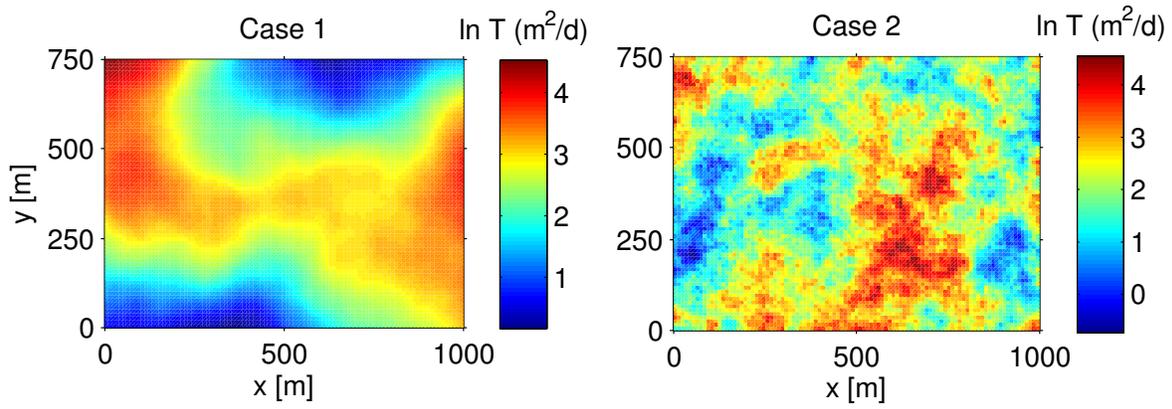
**Figure 1.** Estimation errors ( $\|\mathbf{Q} - \mathbf{Q}_K\|$ ) from 50 randomized eigen-decompositions for the Exponential covariance function  $q(x, x') = \exp\left(-\frac{|x-x'|}{0.3}\right)$  (left) and Gaussian covariance function  $q(x, x') = \exp\left(-\frac{|x-x'|^2}{0.3^2}\right)$  (right). The covariance matrices are constructed on a 100 by 100 grid in  $x \in [0, 1]^2$ . Theoretical minimum errors of the  $K$ -rank approximation ( $k+1$ -th largest eigenvalue  $\lambda_{i+1}$ ) are also plotted.

Method	Textbook	PCGA
number of forward model runs	$n + 1$	$K + p + 2$
matrix multiplication	$\mathcal{O}(mn \log n)$	$\mathcal{O}(mK \log m)$
storage	$\mathcal{O}(mn)$	$\mathcal{O}(mK)$

**Table 1.** Performance comparison of the conventional geostatistical approach and the PCGA. We refer the conventional geostatistical approach with the adjoint-state method and the fast linear algebra method as the “textbook” approach.



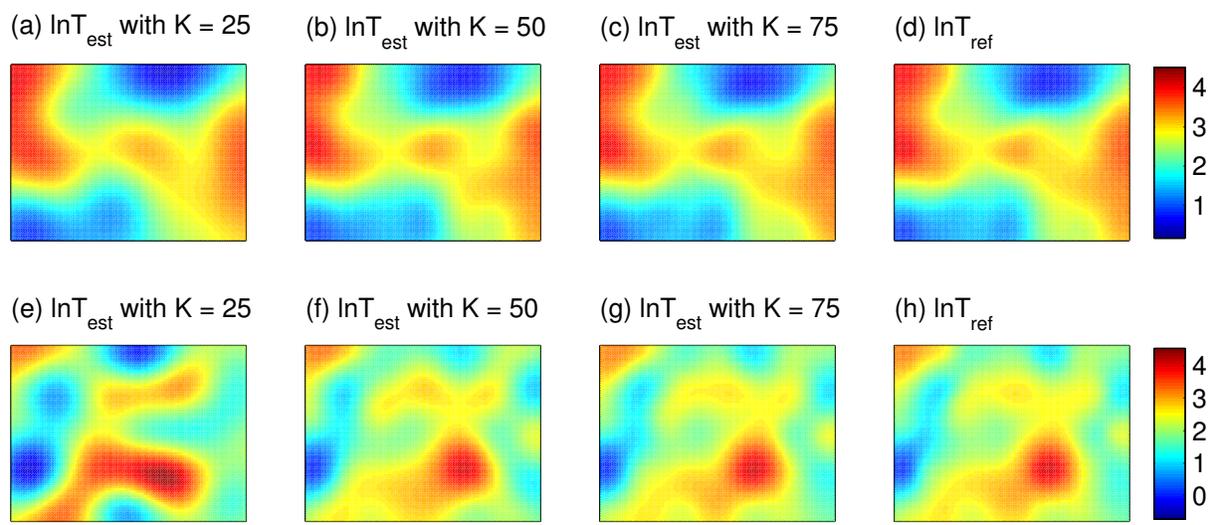
**Figure 2.** Domain of Application 1 and the location of 30 well locations



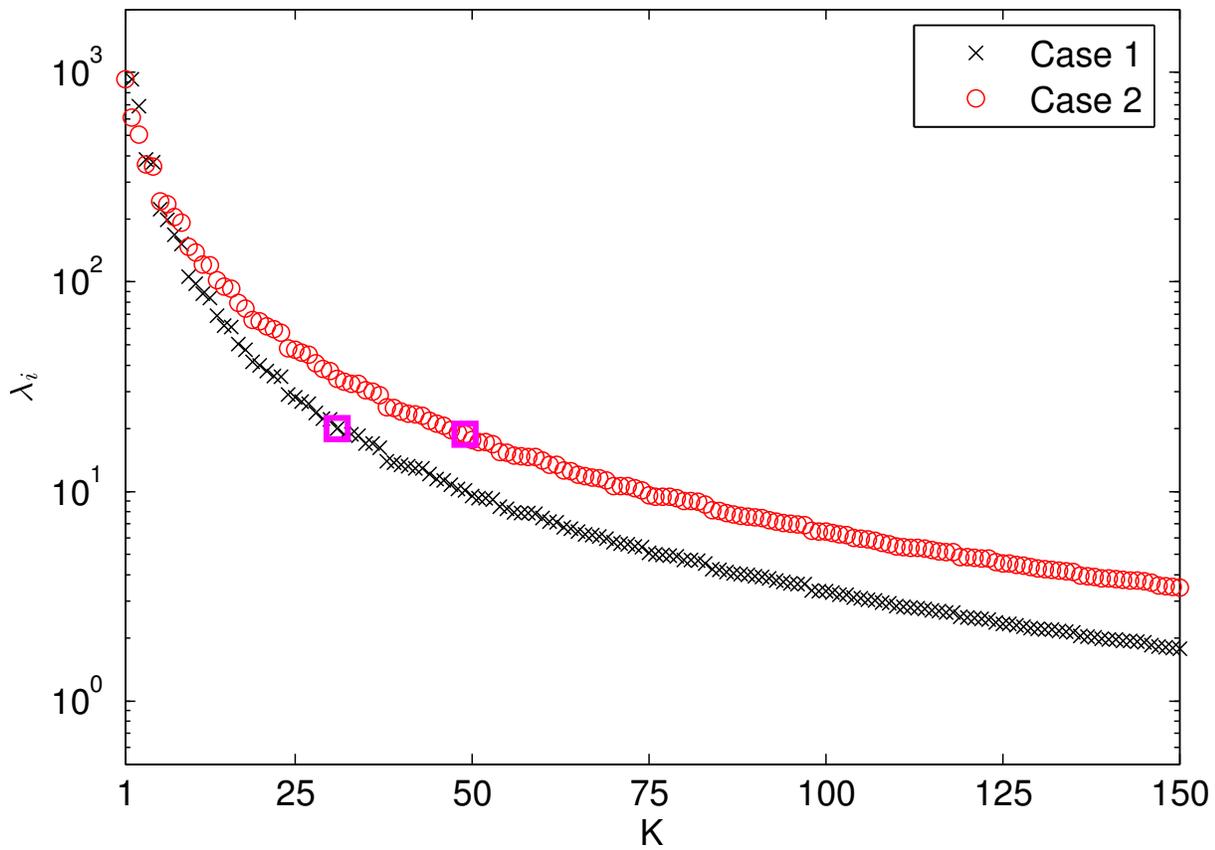
**Figure 3.** True log-transmissivity fields for Case 1 (left) and Case 2 (right): the  $\ln T$  fields were generated from the cubic covariance model  $q(x, x') = (0.002|x - x'|)^3$  and the exponential model  $q(x, x') = \exp(-|x - x'|/150)$

**Table 2.** Parameters used in Application 1

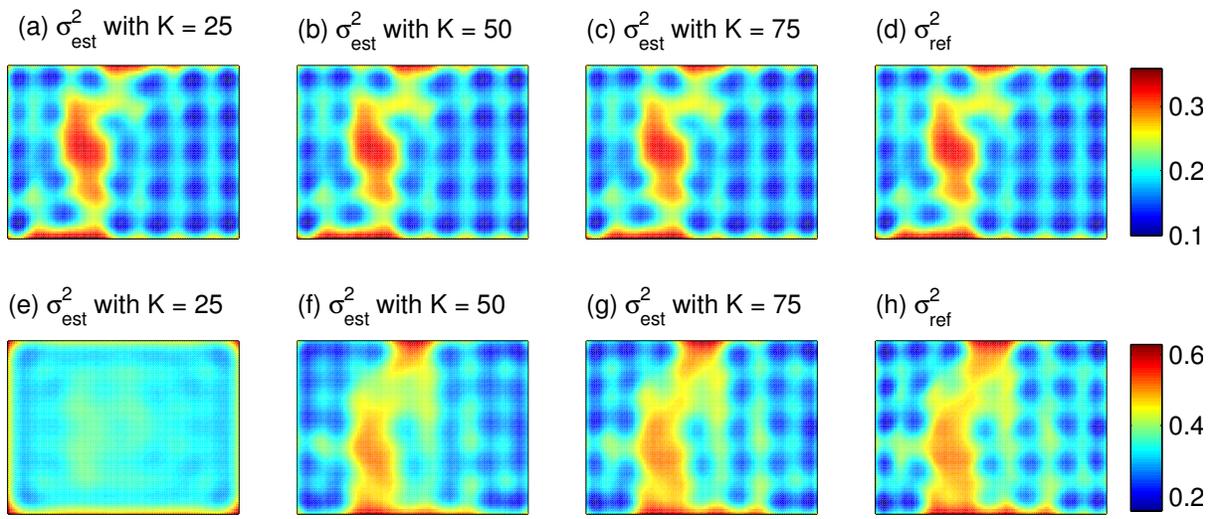
description	Application 1
Geometric Parameters	
$L_x, L_y$ domain length and width (m)	1000, 750
$Q_i$ pumping rates ( $m^2/d$ )	25
$N$ recharge rates ( $m/d$ )	0.001
$\Delta x, \Delta y$ grid size (m)	10
Geostatistical Parameters	
$\ln \bar{T}$ mean log-transmissivity ( $\ln m^2/d$ )	2.5 ( $T = 12.18 m^2/d$ )
$q(x, x')$ covariance kernel for Case 1	$q(x, x') = (0.002 x - x' )^3$
$q(x, x')$ covariance kernel for Case 2	$q(x, x') = \exp(- x - x' /150)$
Measurement Error	
$n_{obs}$ number of measurements	870
$\sigma_h$ standard deviation of the measurement error (m)	0.5



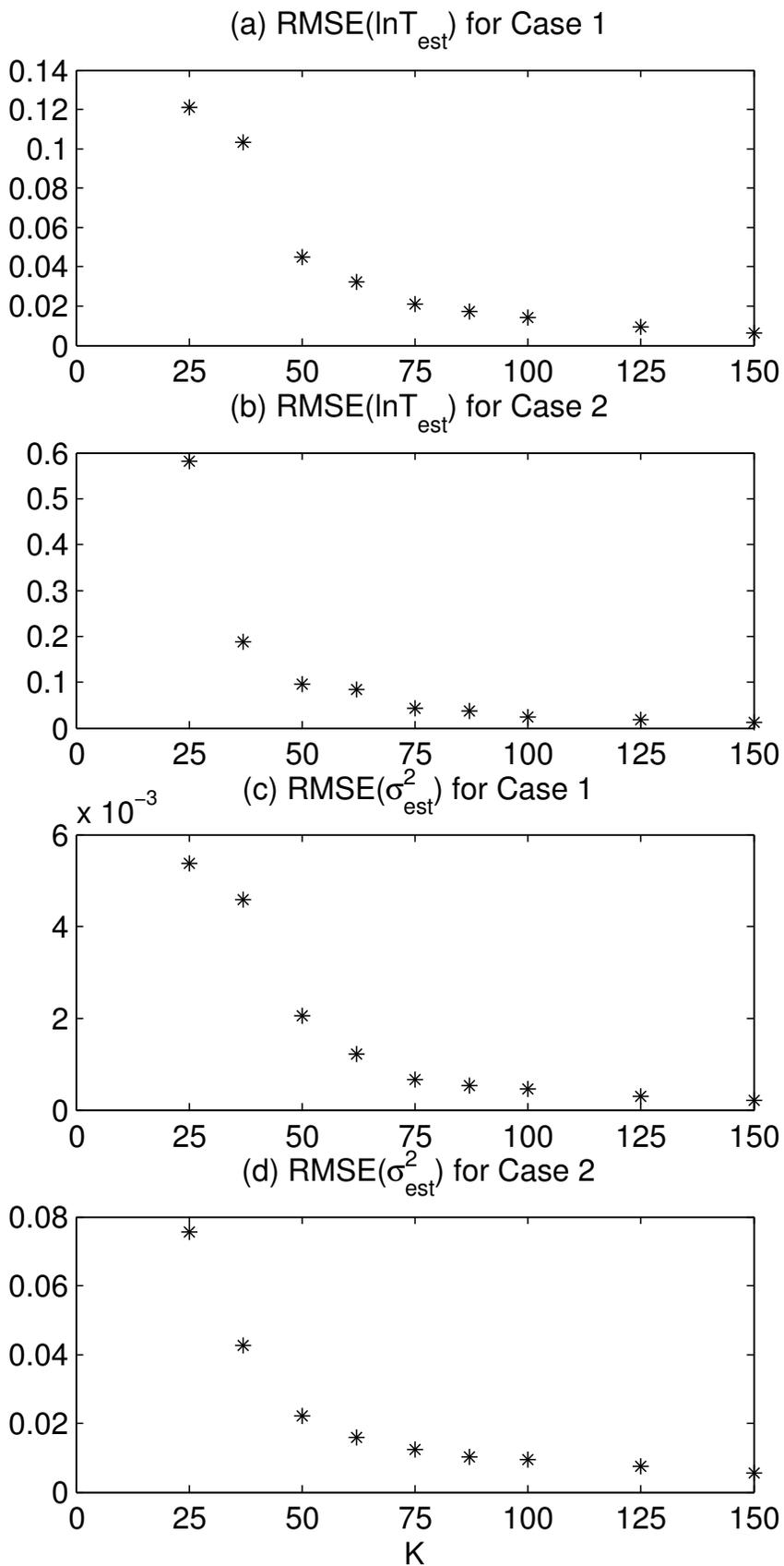
**Figure 4.** The best estimates from the PCGA with  $K = 25, 50$  and  $75$  for Case 1 (a) - (c) and for Case 2 (e) - (g). The best estimates from textbook approach for Case 1 (d) and Case 2 (h) is also plotted as the “reference” estimates.



**Figure 5.** The eigenvalue spectrum of the prior covariance for Case 1 (black) and Case 2 (red);  $\square$  indicates the 31-th (Case 1) and the 49-th (Case 2) eigenvalues whose relative error ( $\lambda_{i+1}/\lambda_i$ ; Eq. (22))  $\approx 0.01$

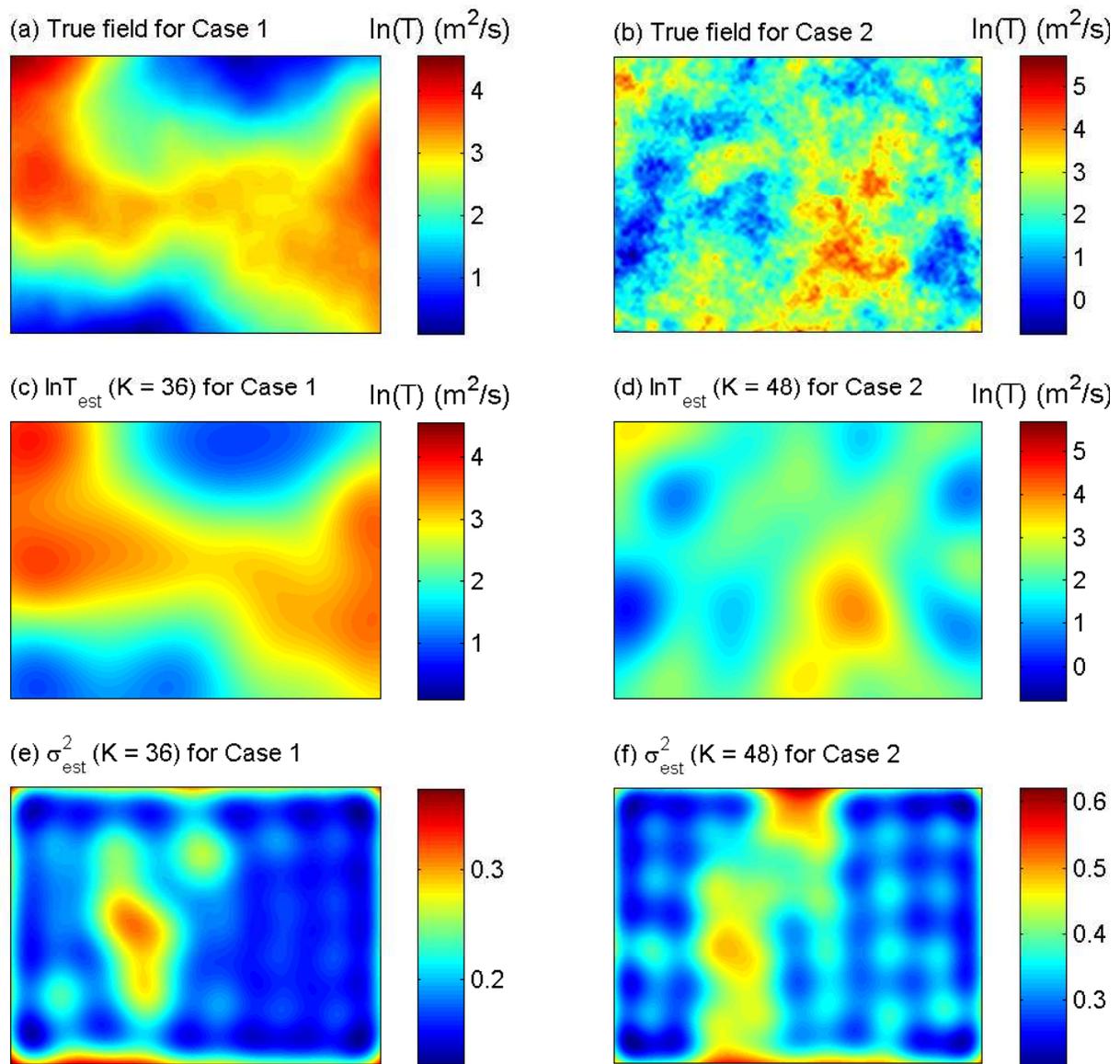


**Figure 6.** The estimation variance map  $\sigma^2(\ln T_{est})$  from the PCGA with  $K = 25, 50$  and  $75$  for Case 1 (a) - (c) and for Case 2 (e) - (g) and the estimation variance map from textbook approach for Case 1 (d) and Case 2 (h)

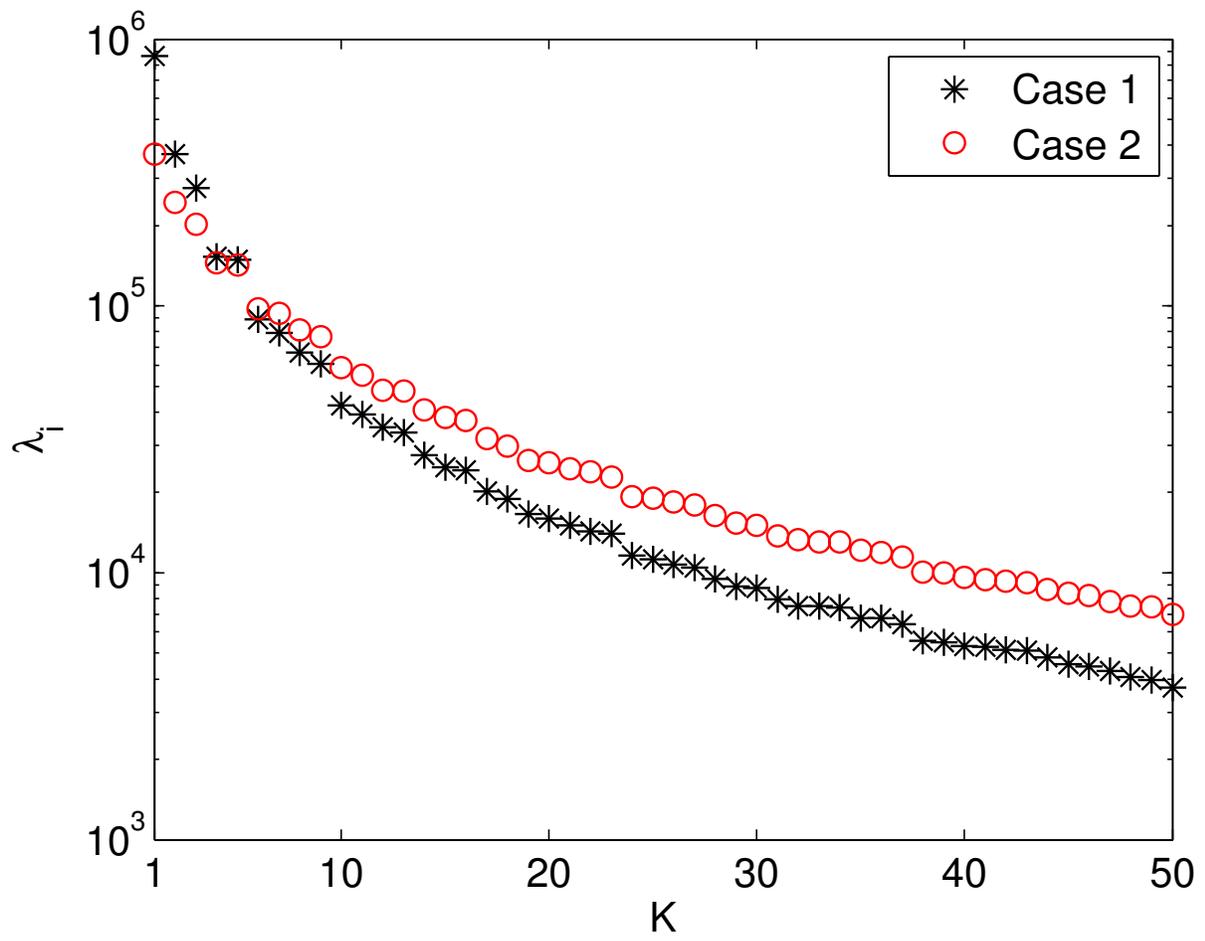


**Figure 7.** The root-mean-square error of the estimates for (a) Case 1 and (b) Case 2

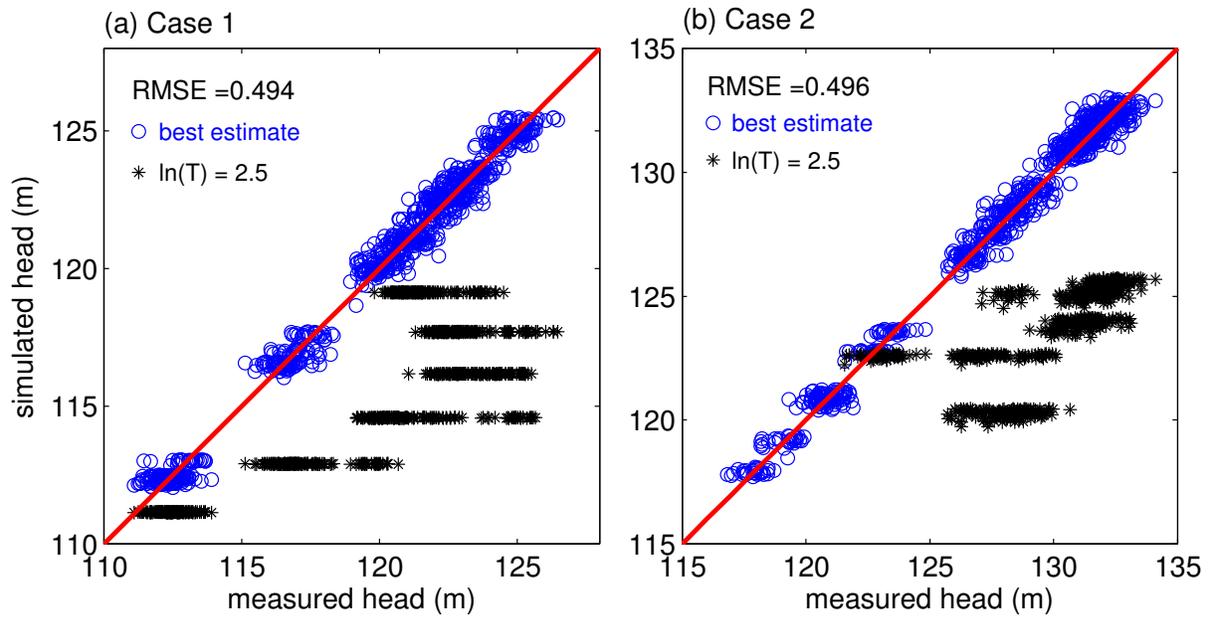
and the estimation variance for (c) Case 1 and (d) Case 2



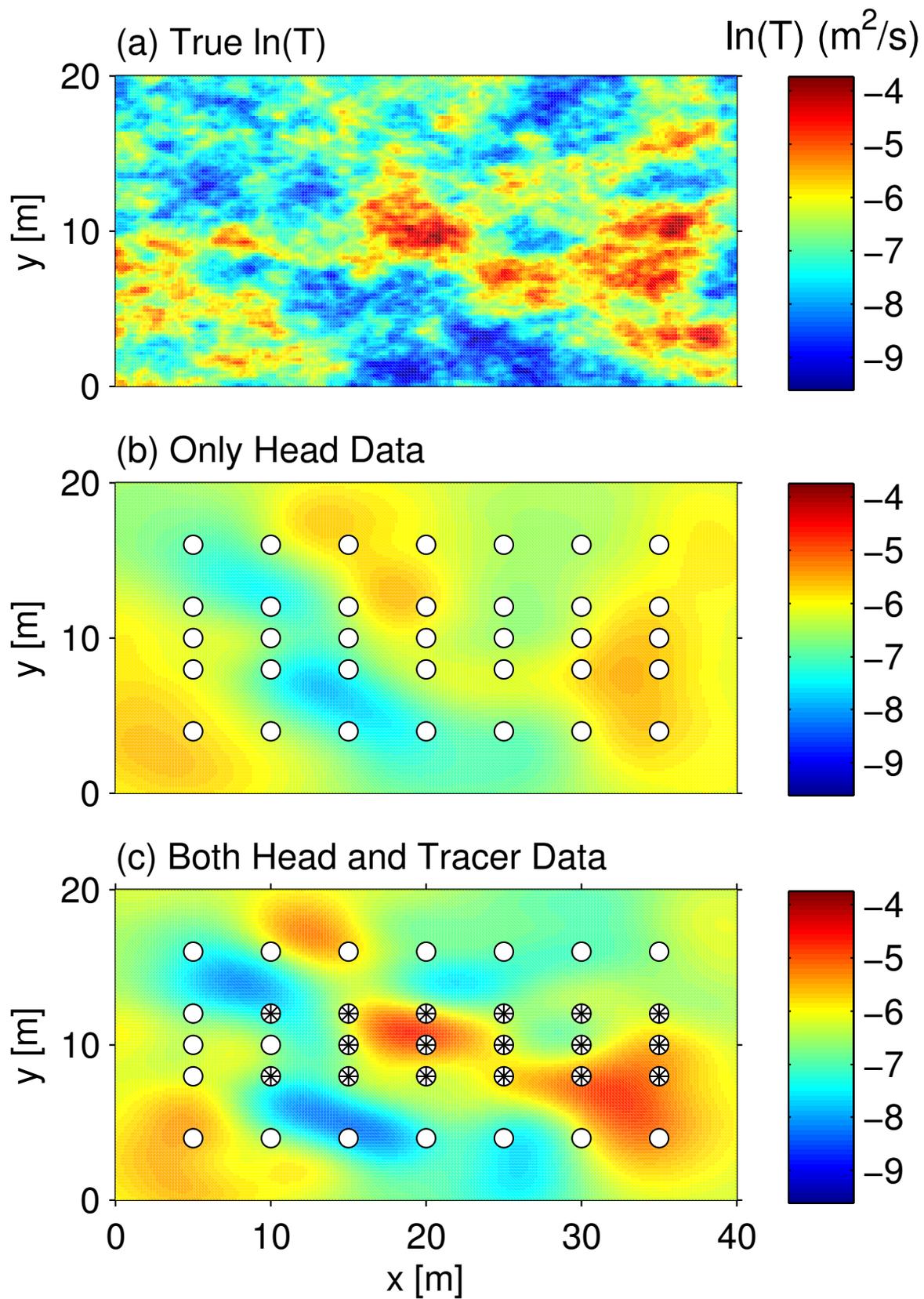
**Figure 8.** True log-transmissivity fields for (a) Case 1 and (b) Case 2; the best estimates for (c) Case 1 and (d) Case 2; corresponding estimation variance for (e) Case 1 and (f) Case 2



**Figure 9.** 50 leading eigenvalues of the  $3 \times 10^6$  by  $3 \times 10^6$  prior covariance matrix used in the Application 2 for (a) Case 1 and (b) Case 2



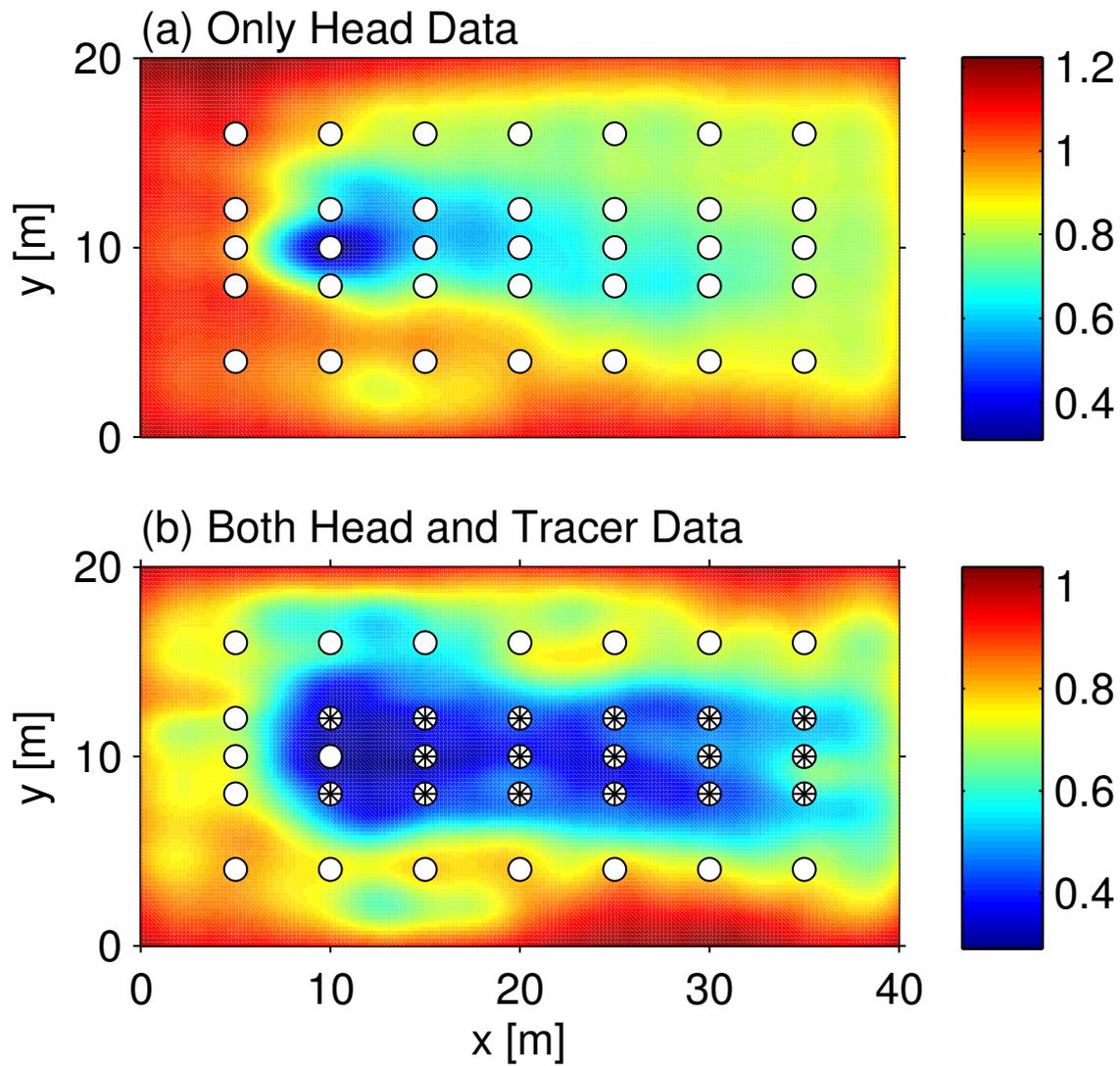
**Figure 10.** Measurement data fitting : measured heads vs. hydraulic heads simulated from the best estimate (blue  $\circ$ ) and heads simulated from the mean log-transmissivity field  $\ln(\bar{T}) = 2.5$  (black  $*$ ) for (a) Case 1 and (b) Case 2



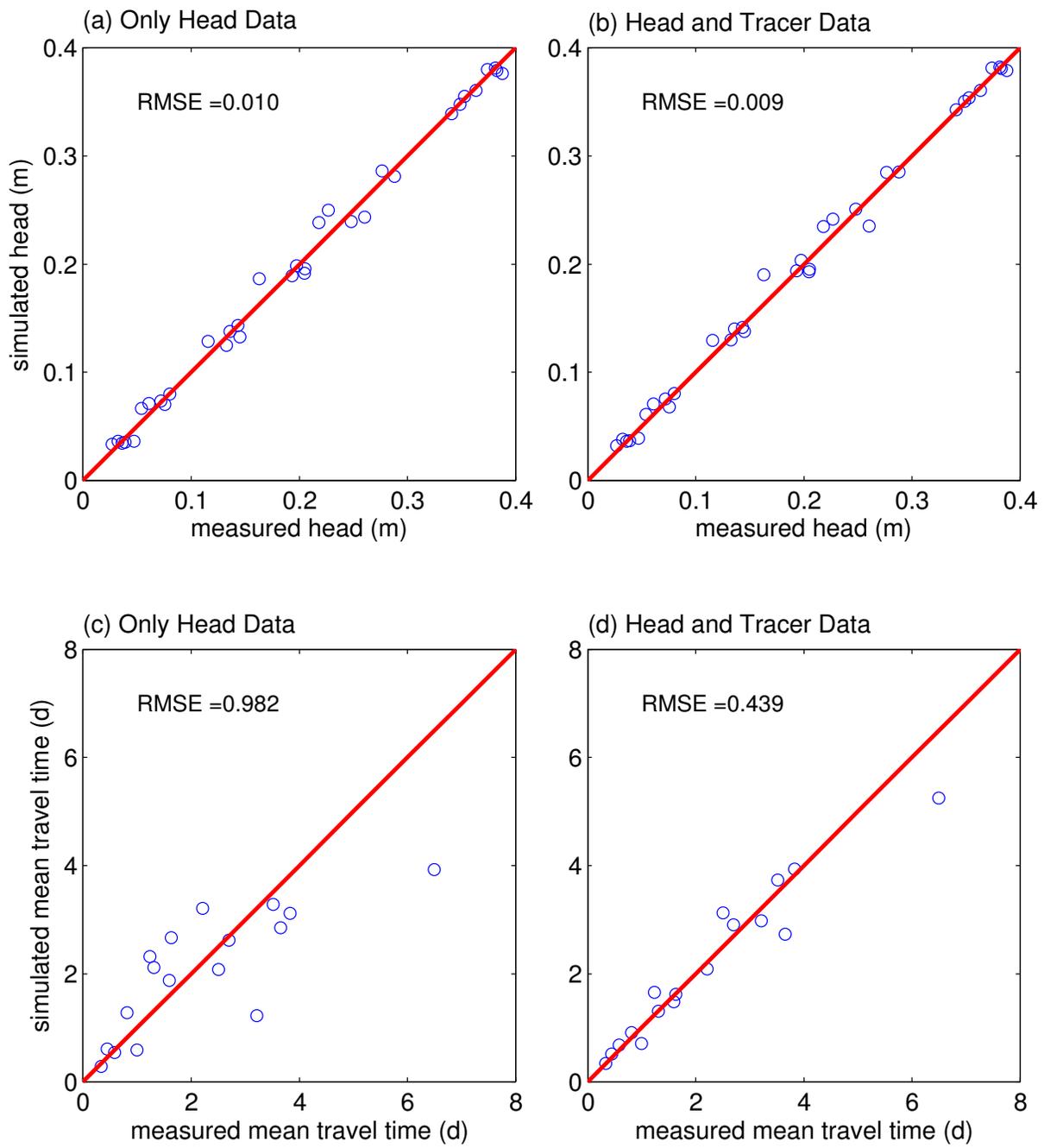
**Figure 11.** (a) “True” log-transmissivity field generated from the exponential model  
D R A F T May 28, 2014, 2:38am D R A F T  
 $q(x, x') = \exp(-|x - x'|/l)$  where  $l_x = 4$  m and  $l_y = 2$  m; (b) the best estimate for head  
data inversion; (c) the best estimate for head and mean travel time data inversion; the  
monitoring locations indicated by  $\circ$  for head and  $*$  for tracer concentration data

**Table 3.** Parameters used in Application 3

Parameter	Description	Value
$L_x, L_y$	domain length and width (m)	40, 20
$\Delta x, \Delta y$	grid spacing (m)	0.2
$h_{x=0}, h_{x=40}$	boundary condition (m)	0.4, 0
$x_w, y_w$	injection well location (m)	10, 10
$Q_w$	injection rate ( $m^2/s$ )	$1.5 \times 10^{-4}$
$\theta$	porosity	0.25
$\alpha_l$	longitudinal dispersivity (m)	0.05
$\alpha_t$	transverse dispersivity (m)	0.01
$D_m$	diffusion coefficient ( $m^2/s$ )	$10^{-9}$
Geostatistical Parameters		
$\ln \bar{T}$	mean of $\ln T$ (m/s)	-7
$q(x, x')$	covariance kernel	$q(x, x') = \exp(- x - x' /l)$
$l_x, l_y$	correlation length in x and y (m)	4, 2
Measurement Error		
$\sigma_h$	standard deviation of measurement error for head (m)	0.01
$\sigma_{\bar{t}_{x0}}$	standard deviation of measurement error for travel time (s)	$0.1 \bar{t}_{x0}$



**Figure 12.** The estimation variance map  $\sigma^2(\ln T_{est})$  for (a) head data and (b) joint inversion; the monitoring locations indicated by  $\circ$  for head and  $*$  for tracer concentration data



**Figure 13.** Measurement data fitting : simulated vs. measured hydraulic head for (a) head data and (b) joint inversion; simulated vs. measured mean travel time for (c) head data and (d) joint inversion