Water Resources Research

METHOD

10.1029/2022WR032610

# High-Dimensional Groundwater Flow Inverse Modeling by Upscaled Effective Model on Principal Components

**Yue Zhao[1], Quan Guo[1], Chunhui Lu[2], and Jian Luo[1]**

[1]School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, USA, [2]State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing, China

**Key Points:**

- Upscaled principal component inverse approach (UPCIA) achieves dimensionality reduction and reduces forward model computation by a grid-coarsening method
- Downscaling and inverse modeling are integrated by evaluating projections on upscaled principal components to invert high-resolution field
- UPCIA can also be used to examine whether finer resolution is needed and possibly to determine an optimal inverse resolution

**Supporting Information:**

Supporting Information may be found in the online version of this article.

**Correspondence to:**

J. Luo,
jian.luo@ce.gatech.edu

**Citation:**

Zhao, Y., Guo, Q., Lu, C., & Luo, J. (2022). High-dimensional groundwater flow inverse modeling by upscaled effective model on principal components. *Water Resources Research*, *58*, e2022WR032610. https://doi.org/10.1029/2022WR032610

Received 16 APR 2022
Accepted 4 JUL 2022

**Author Contributions:**

**Conceptualization:** Jian Luo
**Formal analysis:** Yue Zhao
**Funding acquisition:** Jian Luo
**Methodology:** Yue Zhao, Jian Luo
**Project Administration:** Jian Luo
**Resources:** Chunhui Lu
**Software:** Yue Zhao
**Supervision:** Jian Luo
**Validation:** Yue Zhao, Quan Guo
**Visualization:** Quan Guo
**Writing – original draft:** Yue Zhao
**Writing – review & editing:** Quan Guo, Chunhui Lu, Jian Luo



**Abstract** The main computational costs of gradient-based inverse methods for high-resolution groundwater flow inverse problems include costly forward model simulations and the large number of such simulations required to determine the Jacobian matrix. We develop an upscaling-based inverse approach, named upscaled principal component inverse approach (UPCIA), which achieves dimensionality reduction and reduces computational cost of forward model simulations by evaluating the Jacobian through upscaled effective models on a coarse-resolution grid constructed from upscaled principal components. UPCIA integrates downscaling into the inverse problem by estimating principal component coefficients based on the coarse-resolution forward model, which are then used to generate high-resolution parameter fields. Various numerical experiments demonstrate the effectiveness and efficiency of UPCIA, including 2-D/3-D high-dimensional steady-state and transient hydraulic tomography with known storativity to estimate multi-Gaussian transmissivity or hydraulic conductivity fields. Results show that the hydraulic head is insensitive to small-scale variability of conductivity, and UPCIA provides high-quality inversion results similar to inverse methods with high-resolution forward model simulations and significantly reduces computation time by orders of magnitude. In addition to supporting the characterization of heterogeneity in sufficient detail, UPCIA can also be used to examine whether finer resolution is necessary and possibly to determine an optimal inverse resolution.

## 1. Introduction

The main computational costs of gradient-based inverse methods for solving high-dimensional groundwater inverse problems usually include (a) the storage and computation of large-dimensional spatial covariance matrices, which regularize the smoothness of the parameter field to be estimated, (b) a large number of iterative forward model implementations to determine the Jacobian matrix, and (c) forward model simulations on high-resolution parameter fields (Kitanidis, 2015). Many research efforts have been devoted to improving the storage and computation of large matrices and reducing the number of forward model runs. For example, adjoint state methods were developed to evaluate the Jacobian by solving an additional forward model system for each measurement (Sun & Yeh, 1990a, 1990b, 1992), reducing the number of forward model runs to the number of measurements. Geostatistical approach (GA) has been improved to compute large-dimensional covariance matrices and reduce the number of forward model runs (Ambikasaran et al., 2013; Kitanidis & Lee, 2014; Lee & Kitanidis, 2014; Lin et al., 2017; Liu & Kitanidis, 2011; Nowak & Cirpka, 2004; Nowak et al., 2003; Zha et al., 2018). Kitanidis and Lee (2014) introduced a low-rank approximation of the covariance matrix through principal component analysis for dimensionality reduction, which reduced the number of forward model runs to the retained rank of principal components. Zhao and Luo (2020) reformulated the Bayesian geostatistical inverse approach based on the dimensionality reduction of dominant principal axes, where the inverse problem is transformed from directly estimating the underlying parameter field (Kitanidis & Lee, 2014; Zha et al., 2018) to estimating coefficients or projections on principal axes. This approach is further extended to use an approximate Jacobian via a quasi-Newton method (Zhao & Luo, 2021b) and to account for biased prior structural parameters of spatial covariance by iteratively corrected principal axes (Zhao & Luo, 2021a). In addition, dimensionality reduction can also be achieved by the active subspace method (Yan et al., 2021) and training neural networks on spatial field training images with multipoint geostatistical distribution patterns (Laloy et al., 2018).

All the efforts mentioned above rely on forward model simulations on high-resolution parameter fields. That is, even if the number of forward model runs is reduced (Lee & Kitanidis, 2014; Zhao and Luo, 2020, 2021b), high computing power such as parallelization is still needed to solve the inverse problem in a reasonable time frame, especially for transient forward model simulations such as transient pumping tests (Tiedeman & Barrash, 2020).

So far, we have seen two methods developed to accelerate forward model computation in groundwater inverse problems: the method of temporal moments and surrogate models. The former method reduced the computational cost by transforming transient groundwater flow models to steady-state moment generating equations (Li et al., 2005; Nowak & Cirpka, 2006; Pollock & Cirpka, 2008; Yin & Illman, 2009; Zhu & Yeh, 2006). However, the computation of temporal moments, especially high-order moments, requires long sampling times and smooth measurements, which may be challenging in the field, especially for monitoring wells with heavy tail measurements located in low-permeability regions (Tiedeman & Barrash, 2020; Yin & Illman, 2009). The latter method usually seeks approximate solutions by constructing a solution subspace for reduced order models, which still requires extensive forward model computations for each flow scenario, such as pumping at different wells in high-resolution fields given a smoothness regularization (Boyce & Yeh, 2014; Liu et al., 2013). For example, Liu et al. (2013) proposed to develop a solution subspace for reduced order models based on the cross covariance between the data and estimated parameters, where the number of high-resolution forward model runs depends on the number of measurements and may not be constant for biased prior covariance. Recently, machine learning models were used to develop surrogate models (Xiao et al., 2022). However, training data were also generated by running high-resolution models for given geostatistical moments, including both mean and covariance, which are often uncertain in the field.

In this note, we aim to explore the feasibility of implementing numerical computations of forward models on upscaled coarse grids for inverse estimation of high-dimensional groundwater flow inverse problems. Multigrid methods have been proposed to accelerate inverse modeling, such as stochastic sampling methods, by combining low-fidelity model for screening possible solutions and high-fidelity model for high accuracy and convergence (Peherstorfer et al., 2018), and the multiscale adjoint method for evaluating fine-scale sensitivity coefficients (Fu et al., 2011). Our approach is named the Upscaled Principal Component Inverse Approach (UPCIA), which is based on the reformulated framework to estimate principal component coefficients on reduced dimensions (Zhao & Luo, 2020). The goal is to estimate the retained principal component coefficients using upscaled effective models on coarse-resolution parameter fields and then generate the high-resolution parameter field based on the original principal components, thus integrating downscaling and inverse modeling. This idea is motivated by the observations that numerical simulations of hydraulic heads in a pumping test are insensitive to grid resolution (Kitanidis, 2015), particularly for large-time drawdowns that respond to the effective aggregate transmissivity (Sánchez-Vila et al., 1999, 2006).

## 2. Brief Overview of Reformulated Geostatistical Approach

To be complete, the following presents a brief review of the reformulated geostatistical approach (RGA) for estimating the projections or coefficients on dominant principal axes for large-scale spatial fields (Zhao & Luo, 2020). The general observation equation describing the relationship between data or dependent variables and target parameters is given by

$$\mathbf{y} = \mathbf{h}(\mathbf{s}) + \epsilon \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^{n\times 1}$ represents an observation data vector, $\mathbf{s} \in \mathbb{R}^{m\times 1}$ is an unknown variable vector, $\mathbf{h}$ represents the forward model, and $\epsilon \in \mathbb{R}^{n\times 1}$ represents a Gaussian noise vector with a covariance $\mathbf{R}$ (often proportional to the identity matrix). For spatial random processes, the unknown variable $\mathbf{s}$ is typically conceptualized and modeled by a deterministic mean with unknown drift coefficients and a stochastic process

$$\mathbf{s} = \mathbf{X}\beta + \zeta \tag{2}$$

where $\mathbf{X} \in \mathbb{R}^{m\times p}$ represents the drift of mean, $\beta \in \mathbb{R}^{p\times 1}$ represents the unknown coefficient vector of the drift function, $p$ is the number of mean drifts, and $\zeta \in \mathbb{R}^{m\times 1}$ is typically regularized by the first two moments: zero mean and covariance $\mathbf{Q} \in \mathbb{R}^{m\times m}$, which is often modeled by a two-point geostatistical covariance function.

The stochastic part in $\mathbf{s}$ can be approximated by a linear combination of scaled principal components

$$\mathbf{s} = \mathbf{X}\beta + \mathbf{Z}\alpha \tag{3}$$

where $\mathbf{Z} \in \mathbb{R}^{m\times k}$ is the matrix with $k$ scaled, most dominant eigenvectors or principal components from the covariance matrix of the random process, $\mathbf{Q}_{ss} \in \mathbb{R}^{m\times m}$, satisfying $\mathbf{Q}_{ss} \approx \mathbf{Z}\mathbf{Z}^{\mathrm{T}}$. The latent variable $\alpha \in \mathbb{R}^{k\times 1} \sim \mathrm{N}(0, \mathbf{I})$ is the coefficient vector representing the projections of the stochastic fluctuations on the retained principal axes. The original inverse problem is thus converted to be a lower-dimensional problem of inverting $\alpha$ and $\beta$.

The observation function can then be rewritten as follows:

$$\mathbf{y} = \mathbf{h}(\alpha, \beta) + \epsilon \tag{4}$$

Following the classic Bayesian geostatistical approach (Kitanidis, 1995), the posterior distribution of $\alpha$ and $\beta$ is given by the equation (Zhao & Luo, 2021a):

$$p''(\alpha, \beta) \propto \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{h}(\alpha, \beta))^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{h}(\alpha, \beta)) - \frac{1}{2}\alpha^{\mathrm{T}}\alpha\right) \tag{5}$$

The reformulated approach is to solve the following optimization problem:

$$\min_{\alpha,\beta} f(\alpha, \beta) = \frac{1}{2}(\mathbf{y} - \mathbf{h}(\alpha, \beta))^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y} - \mathbf{h}(\alpha, \beta)) + \frac{1}{2}\alpha^{\mathrm{T}}\alpha \tag{6}$$

To obtain the best estimates of $\alpha$ and $\beta$, Equation 6 is iteratively linearized and minimized by solving the following linear equation system:

$$\begin{bmatrix} \mathbf{H}_\alpha^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}_\alpha + \mathbf{I} & \mathbf{H}_\alpha^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}_\beta \\ \mathbf{H}_\beta^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}_\alpha & \mathbf{H}_\beta^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}_\beta \end{bmatrix} \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_\alpha^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{h}(\alpha, \beta) + \mathbf{H}_\alpha\alpha + \mathbf{H}_\beta\beta\right) \\ \mathbf{H}_\beta^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{h}(\alpha, \beta) + \mathbf{H}_\alpha\alpha + \mathbf{H}_\beta\beta\right) \end{bmatrix} \tag{7}$$

where $\mathbf{H}_\alpha \in \mathbb{R}^{n\times k}$ represents the Jacobian matrix of the forward model with respect to $\alpha$ and $\mathbf{H}_\beta \in \mathbb{R}^{n\times p}$ represents the Jacobian matrix of the forward model with respect to $\beta$. An intuitive way of computing the Jacobian matrices is to compute the differences of forward outputs by sequentially perturbing each element in $\alpha$ and $\beta$.

When the optimization converges, the posterior covariance matrix of $\hat{\alpha}$ is

$$\hat{\mathbf{Q}}_{\alpha\alpha} = \left(\overline{\mathbf{H}}_\alpha^{\mathrm{T}}\mathbf{R}^{-1}\overline{\mathbf{H}}_\alpha + \mathbf{I} - \overline{\mathbf{H}}_\alpha^{\mathrm{T}}\mathbf{R}^{-1}\overline{\mathbf{H}}_\beta\left(\overline{\mathbf{H}}_\beta^{\mathrm{T}}\mathbf{R}^{-1}\overline{\mathbf{H}}_\beta\right)^{-1}\overline{\mathbf{H}}_\beta^{\mathrm{T}}\mathbf{R}^{-1}\overline{\mathbf{H}}_\alpha\right)^{-1} \tag{8}$$

The posterior of the parameter field can be approximated as follows:

$$\hat{\mathbf{Q}}_{ss} = \mathbf{Z}\hat{\mathbf{Q}}_{\alpha\alpha}\mathbf{Z}^{\mathrm{T}} \tag{9}$$

For unknown or biased prior information, the posterior covariance matrix can be used to iteratively correct the prior covariance matrix until the estimation converges (Zhao & Luo, 2021a).

## 3. Upscaled Principal Component Inverse Approach (UPCIA)

RGA admits an efficient and scalable way of solving a Bayesian inverse problem. However, the iterative process consists of repeated formulation of Equation 7, requiring forward model simulations on the high-resolution parameter field. To alleviate the computational burden of running forward models, we propose the UPCIA approach that exploits the connection between high-resolution fields and upscaled coarse-resolution fields. The proposed approach can be regarded as a natural extension of the reformulated geostatistical approach, as the connection between fields of different resolutions can be conveniently established by principal component coefficients. Hereafter, we will denote the low-resolution field as $\mathbf{s}_l$ and distinguish associated variables with subscript "$l$." By applying an upscaling matrix $\mathbf{U}$ to the original field in Equation 3, the low-resolution field can be expressed as follows:

$$\mathbf{s}_l = \mathbf{U}\mathbf{s} = \mathbf{U}\mathbf{X}\beta + \mathbf{U}\mathbf{Z}\alpha = \mathbf{X}_l\beta + \mathbf{Z}_l\alpha \tag{10}$$

where $\mathbf{s}_l \in \mathbb{R}^{m_l\times 1}$ and $m_l$ is a much smaller number than $m$, $\mathbf{U}$ is the upscaling matrix, $\mathbf{X}_l = \mathbf{U}\mathbf{X}$ is the upscaled drift, and $\mathbf{Z}_l = \mathbf{U}\mathbf{Z}$ is the upscaled principal axes. The most common upscaling method is spatial averaging,

which yields the geometric mean of a hydraulic conductivity or transmissivity field. Sanchez-Vila et al. (2006) conducted an extensive review of representative hydraulic conductivities of Gaussian heterogeneous conductivity fields regularized by two-point stationary, isotropic spatial covariance and suggested that traditional hydraulic test interpretations yield conductivities that are closely related to equivalent and/or effective hydraulic conductivities. The upscaled field is a local geometric mean distribution, which still describes large-scale spatial distribution patterns at the coarse resolution.

The moments of the upscaled field can be evaluated

$$E\left[\mathbf{s}_l\right] = \mathbf{X}_l\beta \tag{11}$$

$$E\left[\left(\mathbf{s}_l - \mathbf{X}_l\beta\right)\left(\mathbf{s}_l - \mathbf{X}_l\beta\right)^{\mathrm{T}}\right] = \mathbf{Z}_l\mathbf{Z}_l^{\mathrm{T}} = \mathbf{U}\mathbf{Q}\mathbf{U}^{\mathrm{T}} \tag{12}$$

We shall notice that the upscaled principal axes, $\mathbf{Z}_l$, are not the actual principal axes for the random process on the coarse-resolution grid. First, the upscaled vectors are not necessarily orthogonal to each other and second, the covariance of the upscaled parameter field is not following the covariance model defined for the prior covariance $\mathbf{Q}$ on the high-resolution field. However, the upscaling matrix applies no transformation to the unknown variables; hence, the best estimate of $\alpha$ and $\beta$ can be obtained by using $\mathbf{s}_l$ in the forward model simulation. The observation equation needs to be slightly modified to include additional errors introduced by the low-resolution forward model.

$$\mathbf{y} = \mathbf{h}_l(\alpha, \beta) + \varepsilon_l \tag{13}$$

where $\mathbf{h}_l$ represents the upscaled, low-resolution forward model and $\varepsilon_l \in \mathbb{R}^{n\times 1}$ represents a noise vector with the covariance $\mathbf{T}$. In theory, $\varepsilon_l$ should include the error in Equation 4 and the numerical differences between the high-resolution model, $\mathbf{h}(\alpha, \beta)$, and low-resolution model, $\mathbf{h}_l(\alpha, \beta)$. However, it is challenging to distinguish them. For simplicity, this error covariance matrix $\mathbf{T}$ is modeled similarly to $\mathbf{R}$ as a matrix proportional to the identity matrix as well. Though the Gaussian error assumption does not fully align with the real error distribution, its applicability will be validated in our numerical experiments. Following the same manner, the posterior distribution of $\alpha$ and $\beta$ is derived as

$$p''(\alpha, \beta) \propto \exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{h}_l(\alpha, \beta))^{\mathrm{T}}\mathbf{T}^{-1}\left(\mathbf{y} - \mathbf{h}_l(\alpha, \beta)\right) - \frac{1}{2}\alpha^{\mathrm{T}}\alpha\right) \tag{14}$$

The objective function to be minimized is

$$\min_{\alpha,\beta} f(\alpha, \beta) = \frac{1}{2}(\mathbf{y} - \mathbf{h}_l(\alpha, \beta))^{\mathrm{T}}\mathbf{T}^{-1}\left(\mathbf{y} - \mathbf{h}_l(\alpha, \beta)\right) + \frac{1}{2}\alpha^{\mathrm{T}}\alpha \tag{15}$$

Accordingly, Equation 7 becomes

$$\begin{bmatrix} \mathbf{H}_{\alpha l}^{\mathrm{T}}\mathbf{T}^{-1}\mathbf{H}_{\alpha l} + \mathbf{I} & \mathbf{H}_{\alpha l}^{\mathrm{T}}\mathbf{T}^{-1}\mathbf{H}_{\beta l} \\ \mathbf{H}_{\beta l}^{\mathrm{T}}\mathbf{T}^{-1}\mathbf{H}_{\alpha l} & \mathbf{H}_{\beta l}^{\mathrm{T}}\mathbf{T}^{-1}\mathbf{H}_{\beta l} \end{bmatrix}\begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{\alpha l}^{\mathrm{T}}\mathbf{T}^{-1}\left(\mathbf{y} - \mathbf{h}_l(\alpha, \beta) + \mathbf{H}_{\alpha l}\alpha + \mathbf{H}_{\beta l}\beta\right) \\ \mathbf{H}_{\beta l}^{\mathrm{T}}\mathbf{T}^{-1}\left(\mathbf{y} - \mathbf{h}_l(\alpha, \beta) + \mathbf{H}_{\alpha l}\alpha + \mathbf{H}_{\beta l}\beta\right) \end{bmatrix} \tag{16}$$

where $\mathbf{H}_{\alpha_l}$ and $\mathbf{H}_{\beta_l}$ represent the Jacobian matrices of $\mathbf{h}_l$ with respect to $\alpha$ and $\beta$, respectively.

When the optimization converges, the best estimates of $\hat{\alpha}$ and $\hat{\beta}$ can then be substituted into Equation 3 to obtain the best estimates of the high-resolution field. Also, the posterior covariance matrix of $\hat{\alpha}$ can be evaluated through the upscaled model by

$$\hat{\mathbf{Q}}_{\alpha\alpha} = \left(\mathbf{H}_{\alpha l}^{\mathrm{T}}\mathbf{T}^{-1}\mathbf{H}_{\alpha l} + \mathbf{I} - \mathbf{H}_{\alpha l}^{\mathrm{T}}\mathbf{T}^{-1}\mathbf{H}_{\beta l}\left(\mathbf{H}_{\beta l}^{\mathrm{T}}\mathbf{T}^{-1}\mathbf{H}_{\beta l}\right)^{-1}\mathbf{H}_{\beta l}^{\mathrm{T}}\mathbf{T}^{-1}\mathbf{H}_{\alpha l}\right)^{-1} \tag{17}$$

Equation 9 can then be used to evaluate the variance associated with the best estimates of the high-resolution field.

During the inversion process, the high-resolution field is mainly used to generate its dominant principal axes, $\mathbf{Z}$, given the prior information of spatial covariance. This step can be conveniently implemented by generating

random realizations and singular value decomposition (Kitanidis & Lee, 2014). After the upscaled principal axes, $\mathbf{Z}_l$, are obtained, the high-resolution field $\mathbf{s}$ is not necessarily updated during iterations because the updates of the coarse resolution $\mathbf{s}_l$ can be conveniently computed by Equation 10. Once the optimization routine converges, the high-resolution field $\mathbf{s}$ can be evaluated by Equation 3.

The algorithm of UPCIA is summarized as follows:

1. Evaluate the principal axes, $\mathbf{Z}$, for a given spatial covariance
2. Evaluate the upscaled principal axes, $\mathbf{Z}_l$, for an upscaling matrix, $\mathbf{U}$
3. Estimate $\alpha$ and $\beta$ based on $\mathbf{Z}_l$ in the framework of RGA (Zhao & Luo, 2020)
4. Evaluate the high-resolution field $\mathbf{s}$ and associated uncertainties

UPCIA is developed for given structural parameters for the spatial covariance. Since UPCIA is based on the RGA framework, it is convenient to use the posterior covariance to iteratively correct biased prior structural parameters for unknown spatial covariance (Zhao & Luo, 2021a). The variance of the error term in the matrix $\mathbf{T}$, similar to the covariance matrix of the high-resolution field, $\mathbf{R}$, is unknown a prior. In the numerical implementation, this term can be updated using the mean-squared error between modeling results and measurements for each iteration, which corresponds to a maximum likelihood estimator that can also be achieved by an expectation maximization algorithm (Fienen et al., 2006). However, the error variance decays rapidly with the iterations, that is, data fitting can be rapidly improved. Thus, a minimum threshold should be defined to prevent data fitting from dominating the optimization of the objective function, which may lead to the problem of overfitting (Zhao & Luo, 2021a).

## 4. Numerical Experiments

To test the effectiveness and computational costs of UPCIA, we design various numerical experiments of large-scale hydraulic tomography problems to estimate spatially distributed transmissivity or hydraulic conductivity fields, including 2-D steady-state hydraulic tomography in Gaussian and non-Gaussian fields and highly heterogeneous fields with Gaussian and exponential covariance models, 2-D transient and 3-D steady-state hydraulic tomography.

### 4.1. 2-D Steady-State Hydraulic Tomography

#### 4.1.1. Numerical Experiments and Random Fields

We consider steady-state sequential pumping test in a monitoring well network. The first case is a multi-Gaussian random field with a Gaussian covariance model and the second case is a structured, multimodal, non-Gaussian heterogeneous field. Both transmissivity fields have a fine resolution of $512 \times 512$, which yields 262,144 unknowns to be estimated in the inverse problem. The UPCIA is applied to different upscaling resolutions, including the original $512 \times 512$, $128 \times 128$, $64 \times 64$, $32 \times 32$, and $16 \times 16$. We should notice that UPCIA becomes RGA when the original fine resolution is applied. All numerical experiments are implemented on a desktop computer equipped with Intel® Xeon® W2102 CPU @ 2.90 GHz processor and 8.00 GB RAM.

Figure 1 shows the multi-Gaussian random with a Gaussian covariance model and the non-Gaussian structured field, similar to the porous media packed in a sand box (Liu & Kitanidis, 2011), where several low-permeability inclusions are embedded in a homogeneous medium. The well networks are represented by dots with different colors. Wells used as pumping or monitoring wells alternatively in hydraulic tomography are represented by black dots, while wells only used for pumping are represented by white dots. Hydraulic tomography is implemented by sequentially applying a constant pumping rate at a pumping well and recording steady-state hydraulic heads or drawdowns at all the monitoring wells. For the Gaussian field with 25 wells, we have a total of 600 measurements (25 pumping tests $\times$ 24 monitoring wells) and for the non-Gaussian field with 35 wells, we have an additional 250 measurements (10 additional pumping tests $\times$ 25 monitoring wells). To imitate the actual situation of observational data collected from hydraulic tomography, we also contaminate the synthetic observational data with 5% normal random errors.

Table 1 summarizes the geostatistical and hydraulic parameters of both numerical experiments. It is particularly worth mentioning that because the structured field is not generated by any geostatistical covariance function, the specified geostatistical parameters for regularizing the inversion will be highly biased. Yeh and Liu (2000)

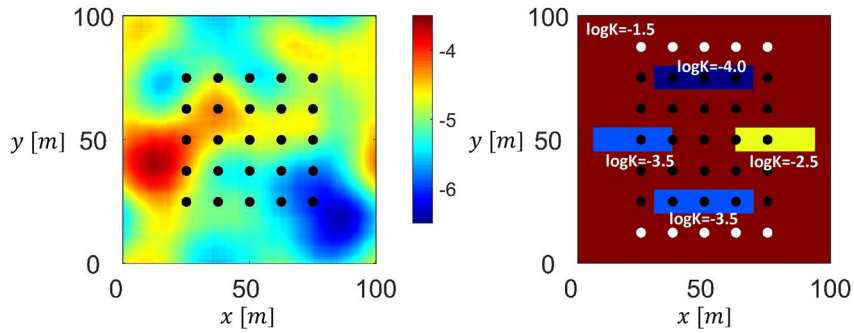


**Figure 1.** Heterogeneous transmissivity fields ($\log_{10} T$) for generating hydraulic tomography data. The left figure is a Gaussian random field with a Gaussian covariance model and the right figure is a non-Gaussian, multimodal, and structured field with low permeability inclusions.

and Liu et al. (2007) discussed that the choice of $\sigma^2$ and $L$ does not make significant differences on the inversion results, particularly in the case when observational data are abundant. Zhao and Luo (2021a) investigated the correction of biased prior structure information. For simplicity, the geostatistical parameters for the logarithmic transmissivity field are assigned as $\sigma^2 = 2.5$, $Lx = 40$ m, $Ly = 20$ m, and $\mu = -1.76$. In both cases, the number of retained principal components are 50, which accounts for more than 95% of the total variance contained in the defined covariance matrix. We keep these parameters constant for all UPCIA simulations.

#### 4.1.2. Results and Discussion

Figure 2 shows the inversion results. The top panel is for the Gaussian random field, while the bottom panel is for structured field. In each panel, the first row shows the best estimates of the logarithmic transmissivity, the second row shows the variance map, and the third row shows the reproduction of measurement. Figure 2 clearly demonstrates that the inversion results from different resolutions are almost the same. The comparison essentially confirms that the UPCIA approach on upscaled coarse grids can capture the main distribution features of the heterogeneous field, both Gaussian and non-Gaussian structured fields, and reproduce the measurements. The variance maps from both cases do not show significant difference among different upscaled resolutions. The lower uncertainty area in the center of the variance map coincides with the location of the well network, indicating that the best estimate is more certain where the data are collected.

Table 2 quantifies the computational performance of the UPCIA on different resolutions, including the number of forward model runs, total computational time, and the map accuracy compared with the true field. For the Gaussian field, we generate multiple fields with the same geostatistical parameters and provide the mean and

**Table 1**
*Geostatistical and Hydrological Parameters for the Synthetic Inverse Problem Cases*

| | | |
|---|---|---|
| Geostatistical properties | | |
| Domain Scale | 100 m × 100 m | 100 m × 100 m |
| Resolution | 512 × 512 | 512 × 512 |
| Covariance Model | Gaussian model | Structured field without any covariance function. |
| Geometric Mean Transmissivity | $T = 10^{-5}$ m$^2$/s | |
| Variance of ln$T$ | $\sigma^2 = 0.5$ | |
| Correlation length | $l_x = l_y = 20$ m | |
| Hydraulic and boundary conditions | | |
| Pumping Rate | $Q = 0.075$ m$^3$/s | $Q = 3.17$ m$^3$/d |
| Top | Impermeable | Impermeable |
| Bottom | Impermeable | Impermeable |
| Left | Drawdown $s = 0$ m | $s = 0$ m |
| Right | $s = 0$ m | $s = 0$ m |

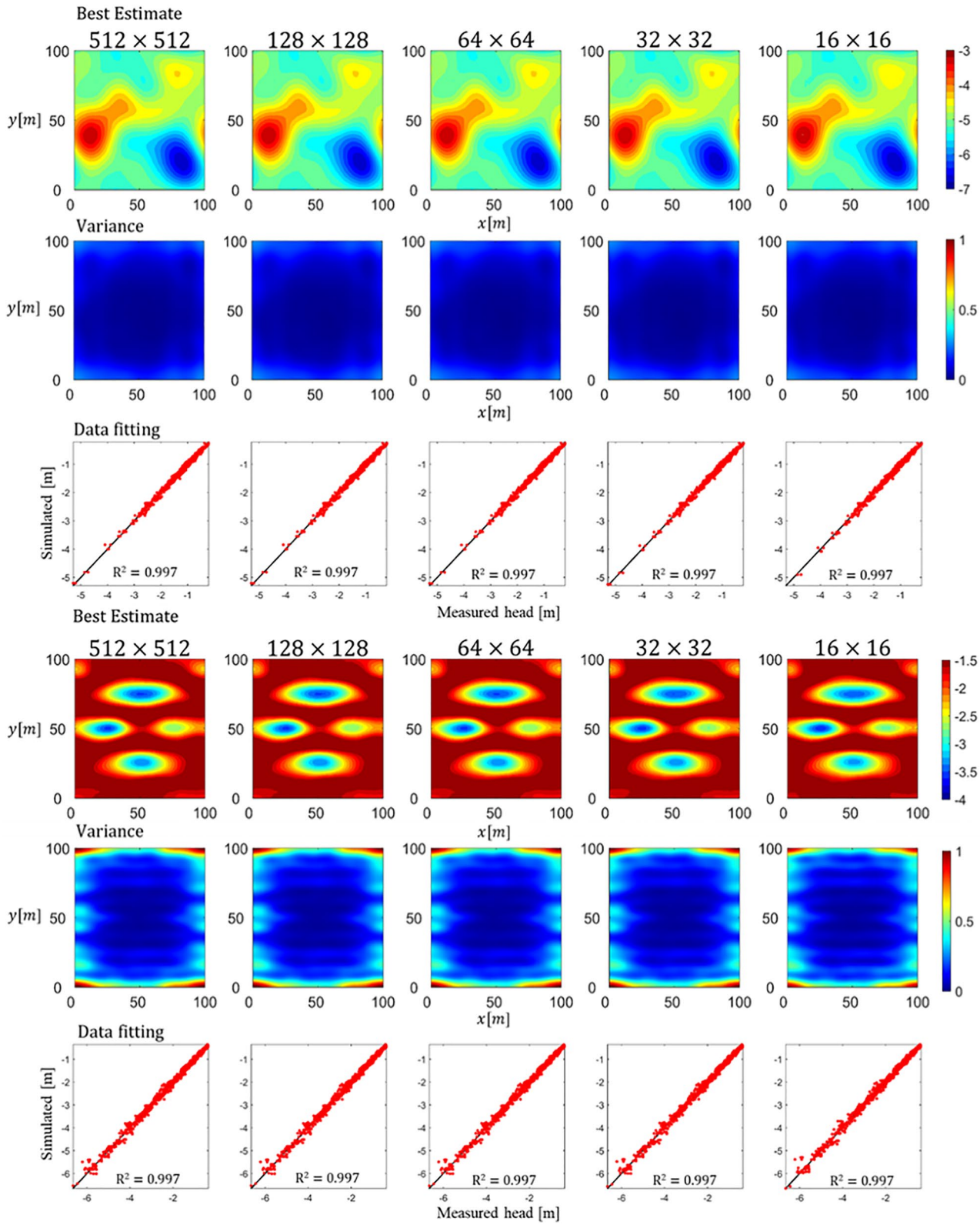


**Figure 2.** Inverse modeling results using upscaled principal component inverse approach with forward model simulations on different resolutions. The top panel is the Gaussian random field and the bottom panel is the structured field. Each panel contains the best estimates (the first row), the variance map for uncertainty quantification (the second row), and the reproduction of hydraulic head measurements (the third row).

**Table 2**
*Computational Performance of Upscaled Principal Component Inverse Approach With Different Upscaled Resolutions for Gaussian and Structured Medium*

| Gaussian/ structured | Number of forward model runs | Total computational time (s) | Map accuracy |
|---|---|---|---|
| $512 \times 512$ | 339 ± 26/371 | 6891 ± 513/10741 | 0.96 ± 0.03/0.65 |
| $128 \times 128$ | 339 ± 26/371 | 307 ± 22/481 | 0.96 ± 0.03/0.65 |
| $64 \times 64$ | 339 ± 26/371 | 61 ± 5/93 | 0.96 ± 0.03/0.65 |
| $32 \times 32$ | 339 ± 26/371 | 10 ± 0.8/15 | 0.96 ± 0.03/0.65 |
| $16 \times 16$ | 339 ± 26/318 | 2.6 ± 0.2/3.2 | 0.96 ± 0.03/0.63 |

*Note*. Multiple Gaussian fields are generated and estimated. The performance results include the mean and confidence interval.

confidence interval of the computational performance to demonstrate that the performance improvement by UPCIA is not case specific. The structured field is only inverted once for performance comparison due to its nonrandomness. The number of forward model runs for the Gaussian field is the same for different forward model resolutions, and also quite constant for the structured field, ranging from 318 to 371, indicating that the number of forward model runs required for convergence does not change or slightly increases with the increase of resolution. Since a more efficient single forward model simulation is performed on a coarser resolution, the total computational time is significantly reduced and the difference from $16 \times 16$ to $512 \times 512$ is more than three orders of magnitude, which is consistent with the computation reduction for linear equation systems. More importantly, there is no substantial deviation in map accuracy at different forward model resolutions, defined by the percentage of grids where the difference between the true and estimated values is less than 10% of the value range (Kang et al., 2017), indicating that the quality of the inversion results using the developed UPCIA is effectively preserved but the computational time is greatly reduced. We also try coarser resolutions, such as $8 \times 8$, but the accuracy is not satisfactory as the forward model becomes inaccurate. To choose an appropriate coarse resolution, one can simply implement UPCIA at different coarse resolutions and compare their performance due to the low computational cost.

The spatial average of the logarithmic hydraulic conductivity values corresponds to the geometric mean of the actual transmissivity values, which is demonstrated to be sufficiently effective for the steady-state hydraulic tomographic problems even under the assumption of Gaussian errors. To quantify the impact of Gaussian error on the computation, we evaluate the difference between the Hessian matrices of the optimization objective function for the upscaled coarse-resolution and original-resolution models during the iteration,

$$\text{Deviation of Hessian matrix} = \frac{\left|\left(\mathbf{H}_{\alpha}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}_{\alpha}+\mathbf{I}\right)-\left(\mathbf{H}_{\alpha l}^{\mathrm{T}}\mathbf{T}^{-1}\mathbf{H}_{\alpha l}+\mathbf{I}\right)\right|_F}{\left|\mathbf{H}_{\alpha}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}_{\alpha}+\mathbf{I}\right|_F}\times 100\% \tag{18}$$

where $||_F$ represents the Frobenius norm. The deviation evaluation method is similar to the comparison of data-covariance variation (Liu et al., 2013). The difference is that our equation system is solved on the retained principal component coefficients, whereas Liu et al. (2013) solved the equation system on the data-covariance space. Figure 3 shows that during the iteration, the deviations between different upscaled resolutions and the original resolution are always kept at low level, thereby having less effect on the equation solution. One of the important reasons is that the forward model output maintains high accuracy across different resolutions under such an upscaling scheme. For forward models with output sensitive to the resolution, special treatment such as non-Gaussian likelihood may be needed (Köpke et al., 2019).

#### 4.1.3. Highly Heterogeneous Fields

A similar steady-state pumping test setup is applied to two highly heterogeneous fields with a variance of six in the log transmissivity, ln$T$. Figure 4 shows the true fields (A1 and B1) and inverse estimation results (A2 and B2). The original resolution is $1{,}024 \times 1{,}024$ with an element size $0.2 \times 0.2$ m. One field has a Gaussian covariance model and the other has an exponential model. We test UPCIA using the upscaled model with a resolution of $32 \times 32$, which solves this high-dimensional inverse problem in minutes. For both cases, the inverse results show high map accuracy, greater than 90%, compared with the true fields (A4 and B4). The Gaussian covariance field is slightly better because its smoothness pattern can be easily maintained at coarse resolution. Data fitting of the steady-state hydraulic head measurements is good for both cases (A5 and B5) and the estimate variances are small (A3 and B3). In particular, the principal component coefficients estimated from the coarse resolutions match the true principal component coefficients at the fine resolution (A6 and B6), particularly for the most dominant principal components, further validating the above results (Figure 3) that the Gaussian error employed in the computation has little effect on evaluating the Hessian matrix for solving the retained principal component coefficients.

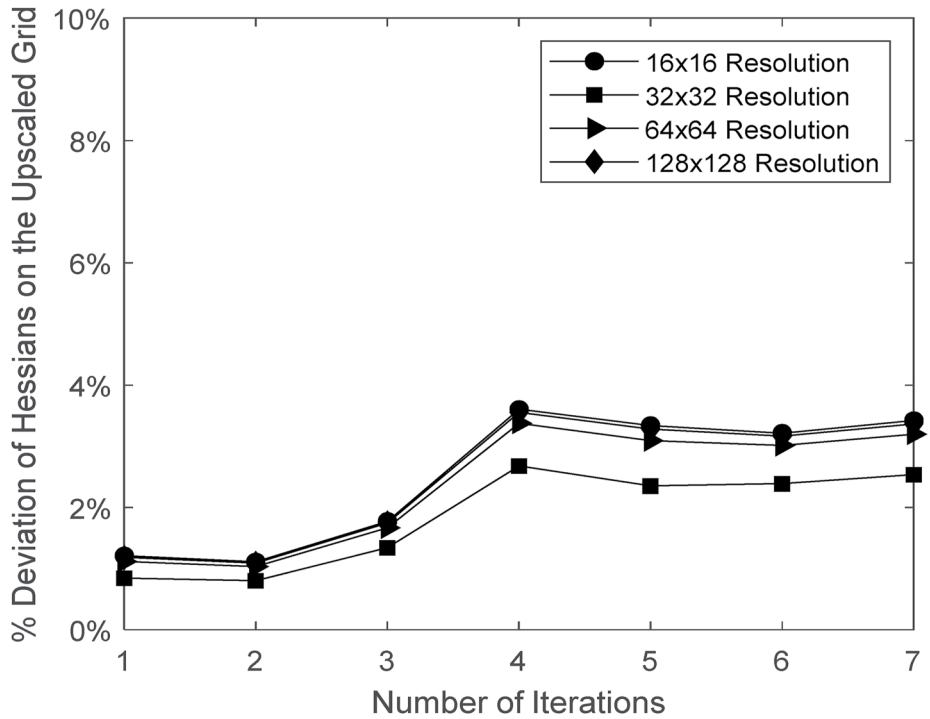


**Figure 3.** Deviations of Hessian matrices in terms of Frobenius norm for different upscaled resolutions of the Gaussian field case.

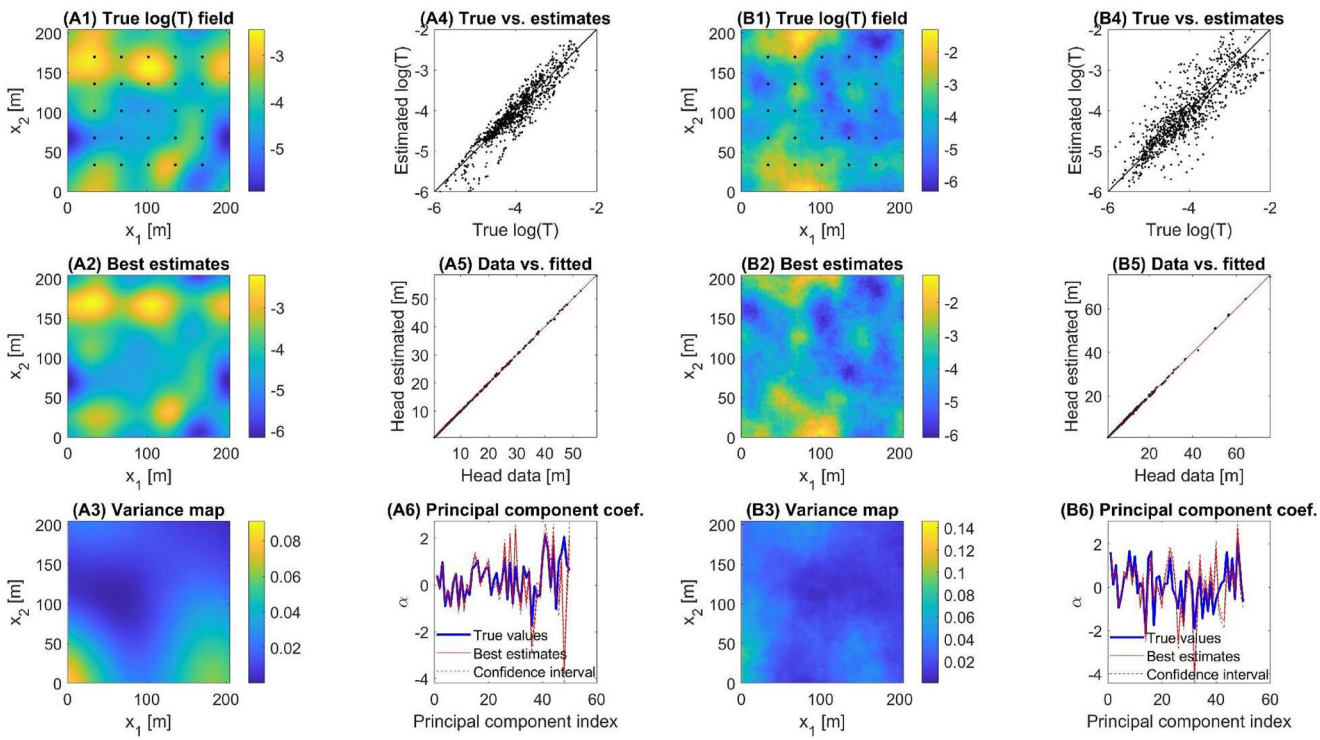


**Figure 4.** Inverse estimation of transmissivity fields at 1,024 × 1,024 resolution from upscaled 32 × 32 coarse resolution models. (a) Gaussian covariance field and (b) exponential covariance field. The black circles in (A1) and (B1) represent the well network. (A2) and (B2) show the best estimates by upscaled principal component inverse approach (UPCIA). (A3) and (B3) show the variance of the best estimates. (A4) and (B4) compare the estimates and the true values. (A5) and (B5) show the measurement fitting. (A6) and (B6) compare the estimated principal component coefficients by UPCIA with the true values.

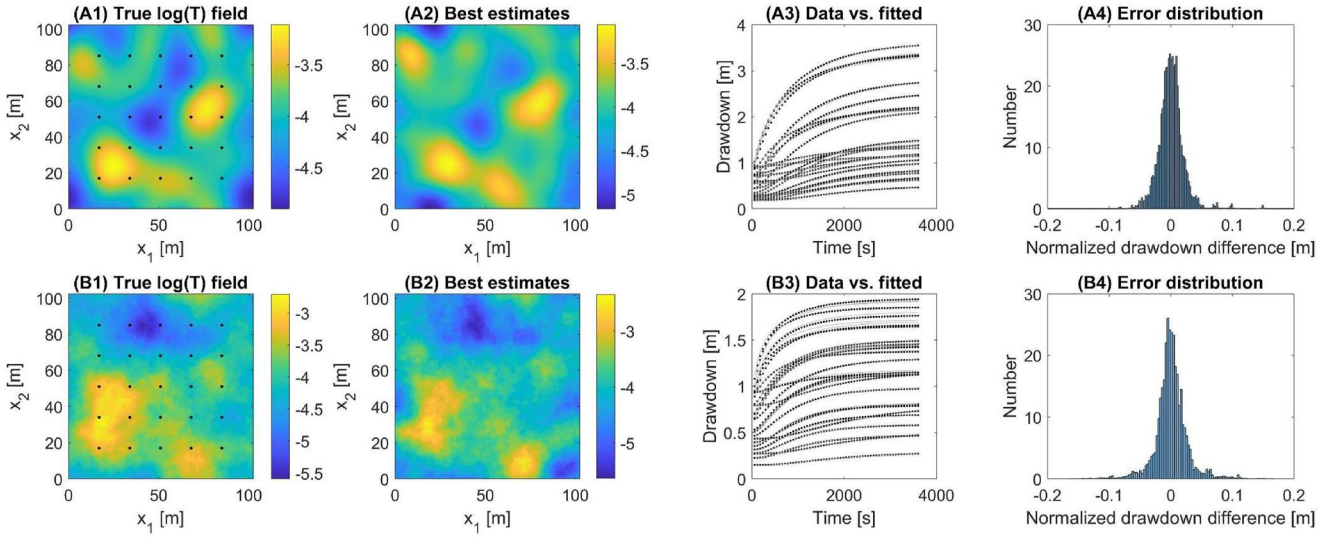


**Figure 5.** Inverse estimation of 2-D transient hydraulic tomography. (a) Gaussian covariance model and (b) exponential covariance model. Black circles in (A1) and (B1) represent the well network. (A2) and (B2) are the inverse estimates from upscaled 32 × 32 coarse resolution models. Black dots in (A3) and (B3) are data measurements and gray lines are model fitting for one pumping test. (A4) and (B4) show the distribution of differences between measurements and coarse-resolution model simulations.

### 4.2. 2-D Transient Hydraulic Tomography

Figure 5 shows the inverse results of transient pumping tests in two Gaussian random fields of log transmissivity, one with a Gaussian covariance model (A1) and the other with an exponential model (B1). Both fields are moderately heterogeneous with a variance of 1. The original resolution is 1,024 × 1,024 with an element size 0.1 × 0.1 m. The storage coefficient for both cases is assumed to be $10^{-4}$ by neglecting the impact of storage heterogeneity (Zhao & Illman, 2021), only the log transmissivity field is inversely estimated. Sequential pumping tests are performed in the well network and 1 hour of time-dependent drawdown data are recorded in one-minute time steps (A3 and B3). Thus, for the well network consisting of 25 wells, the total number of drawdown measurements is 36,000 (25 pumping tests × 24 monitoring wells × 60 measurements). UPCIA uses an upscaled resolution of 32 × 32 and exhibits very good inverse estimates of the log transmissivity fields (A2 and B2). The time-dependent drawdown profiles are also well fitted by the coarse resolution model (A3 and B3). Furthermore, the histograms of the residuals between the measured drawdown and coarse resolution model solutions show Gaussian-type distributions, also confirming that the use of Gaussian error in the upscaled observation function for UPCIA is appropriate.

### 4.3. 3-D Steady-State Hydraulic Tomography

Figure 6 shows the application of UPCIA for estimating a 3-D hydraulic conductivity field based on 3-D steady-state hydraulic tomography. The field resolution is 256 × 256 × 16 with a total of 1.05 million elements and each element size is 1 × 1 × 1 m (Figure 6a). The variance of the Gaussian covariance model is 1 and the correlation length is 40 m horizontally and 6 m vertically. The well network consists of 16 wells, which are uniformly distributed in the domain. The aquifer is confined and all wells are fully penetrated. For each pumping test, multilevel sampling measurements of hydraulic heads are recorded in monitoring wells in each layer (Figure 6b). Thus, the total number of measurements is 3,840 (16 pumping tests × 15 monitoring wells × 16 layers) for inverse modeling. Principal component analysis is performed on the fine-resolution field, and 50 principal components are retained. The upscale principal components have a resolution of 32 × 32 × 16. The estimated log-conductivity field shows good map accuracy compared with the true field (Figure 6c), and the model-predicted hydraulic heads match the measured values (Figure 6d), demonstrating that UPCIA is also applicable to the proposed 3-D case. It is particularly worth mentioning that both the 3-D case and the transient 2-D case are high dimensional and usually require high computing power and long computing time but both can be solved in a few minutes with a PC using UPCIA. In addition, we shall note that for 3-D cases the geometric mean

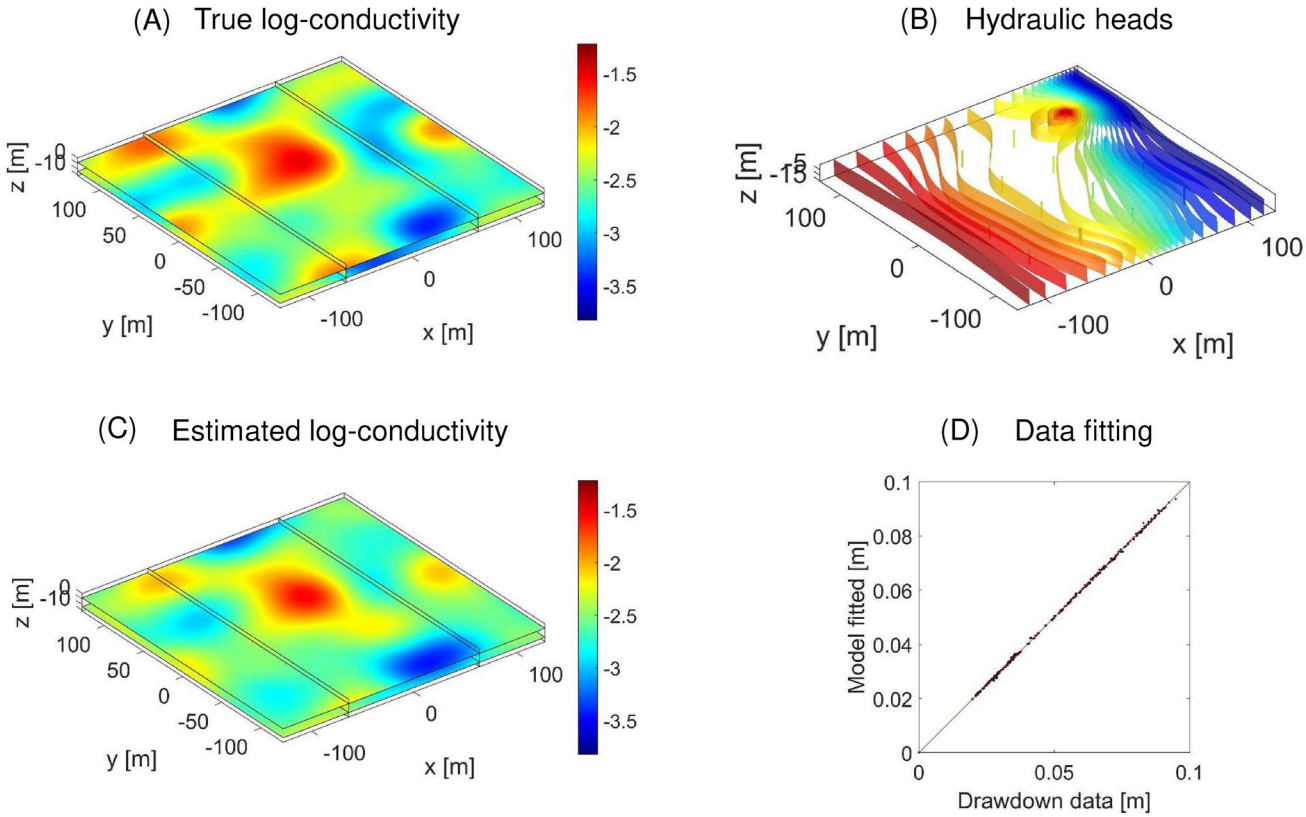


**Figure 6.** Inverse modeling of 3-D steady-state hydraulic tomography for estimating a log hydraulic conductivity field. (a) True field of the 3-D random field with a Gaussian covariance model at a resolution of 256 × 256 × 16. (b) Hydraulic head distributions for a pumping test. (c) upscaled principal component inverse approach estimation based on an upscaled model at a resolution of 32 × 32 × 16. (d) Data fitting.

is usually not the effective hydraulic conductivity for uniform flow in a stationary isotropic conductivity field (Gelhar & Axness, 1983; Gutjahr et al., 1978; Sanchez-Vila et al., 2006). The effective value typically involves additional terms as a function of the variance of log conductivity, which yields an additional constant for the effective log conductivity (De Wit, 1995; Gutjahr et al., 1978; Sanchez-Vila et al., 2006). By including that constant or grouping it into the unknown mean to be estimated, the average upscaling process is still valid. We also note that for a nonuniform well flow field, the equivalent hydraulic conductivity is a function of geostatistics and distance from the well (Bellin et al., 2020; Indelman, 2003; Indelman et al., 1996). Such an upscaling formulation is complex to implement and unnecessary for a distribution of local effective or equivalent hydraulic conductivities.

## 5. Discussion and Conclusion

In this study, we develop an upscaled principal component inverse approach (UPCIA) for high-dimensional groundwater flow inverse problems. UPCIA is developed within the RGA framework on reduced dimensions through principal component analysis. It connects the low-dimensional field and high-dimensional field through upscaled principal components to reduce the computational overhead of forward model simulations. The principal component coefficients are estimated by running the forward model on the upscaled effective coarse-grid field and then used to generate the fine-grid, high-resolution field based on the high-resolution principal components. Uncertainty characterization can be implemented accordingly on the upscaled principal components. The proposed approach is tested in various synthetic numerical experiments of hydraulic tomography to estimate spatially distributed transmissivity or hydraulic conductivity fields. The numerical cases include steady-state pumping in 2-D moderately and highly heterogeneous fields, 2-D transient pumping, 3-D steady-state pumping, and a 2-D steady-state pumping in a structured, non-Gaussian field. The results indicate that UPCIA substantially

reduces computation time while still maintaining high accuracy similar to RGA performing forward model simulations on high-resolution fields.

UPCIA uses the local spatial average of logarithmic hydraulic conductivity or transmissivity values for the upscaled coarse-resolution field, which corresponds to a geometric mean distribution of the hydraulic conductivity or transmissivity field. An iteratively updated Gaussian error is used to quantify the overall error between measurements and upscaled coarse-resolution forward model predictions. Our numerical cases confirm the validity of the Gaussian error assumption and demonstrate that it does not significantly affect the evaluation of the Hessian matrix of the objective function during iterations, resulting in accurate estimates of the principal component coefficients on the upscaled principal components. All our tested cases are Gaussian fields with Gaussian or exponential covariance models. The structured, non-Gaussian field with inclusions embedded in a homogeneous field is actually also described by a Gaussian field with a Gaussian covariance model. The Gaussian assumption is a prerequisite for GA and UPCIA, which is also an assumption to use the upscaled coarse-resolution model with effective conductivity or transmissivity. Our upscaled model is an equivalent model with a distributed local geometric mean, which is different from the ensemble effective model (Sanchez-Vila et al., 2006), and still describes large-scale spatial distribution patterns at the coarse resolution. We have shown that the upscaling method provides accurate calculation of hydraulic heads in the given examples as long as the assumption of multi-Gaussian log conductivity is valid. However, our example with the structured inclusions shows that the inversions miss the hard conductivity contrast, which cannot be captured by the smoothness of the multi-Gaussian regularization.

UPCIA integrates downscaling into the inverse modeling through upscaled principal components. That is, only one inverse problem is solved for the principal component coefficients and the fine-resolution field can then be generated. If one inversely estimates the coarse-resolution field and then downscale it to fine resolution, it becomes two separate inverse problems with inconsistent inverse methods. In particular, downscaling from a coarse-resolution field to a fine-resolution field, such as kriging, may still require high computational costs. Furthermore, one may question whether it is necessary to estimate a fine-resolution field given that fine-scale features hardly affect hydraulic heads, which actually motivates the development of UPCIA. In fact, for problems that are originally planned to resolve fine-scale features, such as joint flow and transport problems (Kitanidis, 2015), UPCIA can be used to support characterizing heterogeneity in sufficient detail, examining whether finer resolution is necessary and possibly determining an optimal resolution. Additionally, UPCIA is developed within the framework of RGA, which relies on the dimensionality reduction of principal components. Theoretically, UPCIA cannot be applied to random fields that may not be well-described by principal components and the multi-Gaussian assumption, such as binary connected fields. In fact, flow and transport behavior in such fields can be sensitive to small-scale heterogeneity (Kitanidis, 2015), which can make it difficult to apply the idea of UPCIA. We are conducting research to examine the effectiveness of UPCIA in such applications. However, regardless of its theoretical limitations, the simplicity and low computational cost of UPCIA always make it an attractive option to practice when fine-resolution forward models are needed, even for gradient-free methods (Park, 2020; Park & Lee, 2021; Yan et al., 2021). We can conveniently apply upscaled effective models at two different coarse resolutions and compare their performance to determine if finer resolution is required.

## Data Availability Statement

Matlab code is provided in the Text S1, while also stored in Github: https://github.com/yuezhao001/upscaled_effective_model. All data used in the study is synthetically generated. No other data is needed.

**Acknowledgments**
The authors would like to thank the editor, associate editor (Dr. Olaf A. Cirpka), and four anonymous reviewers for their constructive comments, which greatly improve the quality of the manuscript. We thank Dr. Cirpka for providing Matlab code for groundwater flow simulation.

## References

Ambikasaran, S., Li, J. Y., Kitanidis, P. K., & Darve, E. (2013). Large-scale stochastic linear inversion using hierarchical matrices. *Computational Geosciences*, *17*(6), 913–927. https://doi.org/10.1007/s10596-013-9364-0

Bellin, A., Fiori, A., & Dagan, G. (2020). Equivalent and effective conductivities of heterogeneous aquifers for steady source flow, with illustration for hydraulic tomography. *Advances in Water Resources*, *142*, 103632. https://doi.org/10.1016/j.advwatres.2020.103632

Boyce, S. E., & Yeh, W. W.-G. (2014). Parameter-independent model reduction of transient groundwater flow models: Application to inverse problems. *Advances in Water Resources*, *69*, 168–180. https://doi.org/10.1016/j.advwatres.2014.04.009

De Wit, A. (1995). Correlation structure dependence of the effective permeability of heterogeneous porous media. *Physics of Fluids*, *7*(11), 2553–2562. https://doi.org/10.1063/1.868705

Fienen, M. N., Luo, J., & Kitanidis, P. K. (2006). A Bayesian geostatistical transfer function approach to tracer test analysis. *Water Resources Research*, *42*(7), W07426. https://doi.org/10.1029/2005wr004576

Fu, J., Caers, J., & Tchelepi, H. A. (2011). A multiscale method for subsurface inverse modeling: Single-phase transient flow. *Advances in Water Resources*, *34*(8), 967–979. https://doi.org/10.1016/j.advwatres.2011.05.001

Gelhar, L. W., & Axness, C. L. (1983). Three-dimensional stochastic-analysis of macrodispersion in aquifers. *Water Resources Research*, *19*(1), 161–180. https://doi.org/10.1029/wr019i001p00161

Gutjahr, A. L., Gelhar, L. W., Bakr, A. A., & MacMillan, J. R. (1978). Stochastic analysis of spatial variability in subsurface flows: 2. Evaluation and application. *Water Resources Research*, *14*(5), 953–959. https://doi.org/10.1029/wr014i005p00953

Indelman, P. (2003). Transient well-type flows in heterogeneous formations. *Water Resources Research*, *39*(3), 1064. https://doi.org/10.1029/2002WR001407

Indelman, P., Fiori, A., & Dagan, G. (1996). Steady flow toward wells in heterogeneous formations: Mean head and equivalent conductivity. *Water Resources Research*, *32*(7), 1975–1983. https://doi.org/10.1029/96wr00990

Kang, P. K., Lee, J., Fu, X., Lee, S., Kitanidis, P. K., & Juanes, R. (2017). Improved characterization of heterogeneous permeability in saline aquifers from transient pressure data during freshwater injection. *Water Resources Research*, *53*(5), 4444–4458. https://doi.org/10.1002/2016wr020089

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*(8), 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*(7), 5428–5443. https://doi.org/10.1002/2013wr014630

Köpke, C., Irving, J., & Roubinet, D. (2019). Stochastic inversion for soil hydraulic parameters in the presence of model error: An example involving ground-penetrating radar monitoring of infiltration. *Journal of Hydrology*, *569*, 829–843. https://doi.org/10.1016/j.jhydrol.2018.12.016

Laloy, E., Herault, R., Jacques, D., & Linde, N. (2018). Training-image based geostatistical inversion using a spatial generative adversarial neural network. *Water Resources Research*, *54*(1), 381–406. https://doi.org/10.1002/2017wr022148

Lee, J., & Kitanidis, P. K. (2014). Large-scale hydraulic tomography and joint inversion of head and tracer data using the principal component geostatistical approach (PCGA). *Water Resources Research*, *50*(7), 5410–5427. https://doi.org/10.1002/2014wr015483

Li, W., Nowak, W., & Cirpka, O. A. (2005). Geostatistical inverse modeling of transient pumping tests using temporal moments of drawdown. *Water Resources Research*, *41*(8), W08403. https://doi.org/10.1029/2004wr003874

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*(8), 6784–6801. https://doi.org/10.1002/2016wr020299

Liu, X., Illman, W. A., Craig, A., Zhu, J., & Yeh, T.-C. J. (2007). Laboratory sandbox validation of transient hydraulic tomography. *Water Resources Research*, *43*(5), 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*(2), 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*(10), 6587–6600. https://doi.org/10.1002/wrcr.20489

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., & Cirpka, O. A. (2006). Geostatistical inference of hydraulic conductivity and dispersivities from hydraulic heads and tracer data. *Water Resources Research*, *42*(8), W08416. https://doi.org/10.1029/2005wr004832

Nowak, W., Tenkleve, S., & Cirpka, O. A. (2003). Efficient computation of linearized cross-covariance and auto-covariance matrices of interdependent quantities. *Mathematical Geology*, *35*(1), 53–66. https://doi.org/10.1023/a:1022365112368

Park, E. (2020). A geostatistical evolution strategy for subsurface characterization: Theory and validation through hypothetical two-dimensional hydraulic conductivity fields. *Water Resources Research*, *56*(3), e2019WR026922. https://doi.org/10.1029/2019wr026922

Park, E., & Lee, J. (2021). A non-Bayesian nonparametric model for characterization of basin-scale aquifers using groundwater level fluctuations. *Journal of Hydrology*, *602*, 126710. https://doi.org/10.1016/j.jhydrol.2021.126710

Peherstorfer, B., Willcox, K., & Gunzburger, M. (2018). Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review*, *60*(3), 550–591. https://doi.org/10.1137/16m1082469

Pollock, D., & Cirpka, O. A. (2008). Temporal moments in geoelectrical monitoring of salt tracer experiments. *Water Resources Research*, *44*(12), W12416. https://doi.org/10.1029/2008wr007014

Sánchez-Vila, X., Guadagnini, A., & Carrera, J. (2006). Representative hydraulic conductivities in saturated groundwater flow. *Reviews of Geophysics*, *44*(3), RG3002. https://doi.org/10.1029/2005rg000169

Sánchez-Vila, X., Meier, P. M., & Carrera, J. (1999). Pumping tests in heterogeneous aquifers: An analytical study of what can be obtained from their interpretation using Jacob's method. *Water Resources Research*, *35*(4), 943–952. https://doi.org/10.1029/1999wr900007

Sun, N. Z., & Yeh, W. W. G. (1990a). Coupled inverse problems in groundwater modeling: 1. Sensitivity analysis and parameter identification. *Water Resources Research*, *26*(10), 2507–2525. https://doi.org/10.1029/wr026i010p02507

Sun, N. Z., & Yeh, W. W. G. (1990b). Coupled inverse problems in groundwater modeling: 2. Identifiability and experimental design. *Water Resources Research*, *26*(10), 2527–2540. https://doi.org/10.1029/wr026i010p02527

Sun, N. Z., & Yeh, W. W. G. (1992). A stochastic inverse solution for transient groundwater flow: Parameter identification and reliability analysis. *Water Resources Research*, *28*(12), 3269–3280. https://doi.org/10.1029/92wr00683

Tiedeman, C. R., & Barrash, W. (2020). Hydraulic tomography: 3D hydraulic conductivity, fracture network, and connectivity in mudstone. *Groundwater*, *58*(2), 238–257. https://doi.org/10.1111/gwat.12915

Xiao, C., Zhang, S., Ma, X., Jin, J., & Zhou, T. (2022). *Model-reduced adjoint-based inversion using deep-learning: Example of geological carbon sequestration modelling*. Water Resources Research, e2021WR031041.

Yan, H., Hao, C., Zhang, J., Illman, W. A., Lin, G., & Zeng, L. (2021). Accelerating groundwater data assimilation with a gradient-free active subspace method. *Water Resources Research*, *57*(12), e2021WR029610. https://doi.org/10.1029/2021wr029610

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

Yin, D., & Illman, W. A. (2009). Hydraulic tomography using temporal moments of drawdown recovery data: A laboratory sandbox study. *Water Resources Research*, *45*(1), W01502. https://doi.org/10.1029/2007wr006623

Zha, Y., Yeh, T. C. J., Illman, W. A., Zeng, W., Zhang, Y., Sun, F., & Shi, L. J. (2018). A reduced-order successive linear estimator for geostatistical inversion and its application in hydraulic tomography. *Water Resources Research*, *54*(3), 1616–1632. https://doi.org/10.1002/2017wr021884

Zhao, Y., & Luo, J. (2020). Reformulation of Bayesian geostatistical approach on principal components. *Water Resources Research*, *56*(4), e2019WR026732. https://doi.org/10.1029/2019wr026732

Zhao, Y., & Luo, J. (2021a). Bayesian inverse modeling of large-scale spatial fields on iteratively corrected principal components. *Advances in Water Resources*, *151*, 103913. https://doi.org/10.1016/j.advwatres.2021.103913

Zhao, Y., & Luo, J. (2021b). A quasi-Newton reformulated geostatistical approach on reduced dimensions for large-dimensional inverse problems. *Water Resources Research*, *57*(1), e2020WR028399. https://doi.org/10.1029/2020wr028399

Zhao, Z., & Illman, W. A. (2021). On the importance of considering specific storage heterogeneity in hydraulic tomography: Laboratory sandbox and synthetic studies. *Journal of Hydrology*, *593*, 125874. https://doi.org/10.1016/j.jhydrol.2020.125874

Zhu, J., & Yeh, T. C. J. (2006). Analysis of hydraulic tomography using temporal moments of drawdown recovery data. *Water Resources Research*, *42*(2), W02403. https://doi.org/10.1029/2005wr004309