AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2014WR016179

Linear functional minimization for inverse modeling

D. A. Barajas-Solano[1], B. E. Wohlberg[2], V. V. Vesselinov[3], and D. M. Tartakovsky[1]

[1]Department of Mechanical and Aerospace Engineering, University of California, San Diego, California, USA, [2]Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA, [3]Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA

Key Points:
- New inverse modeling technique for systems with piecewise continuous parameters
- New linearized functional minimization method for nonlinear parameter estimation
- Method proposed resolves large-scale features of parameters from sparse data

Correspondence to:
D. M. Tartakovsky,
dmt@ucsd.edu

Citation:
Barajas-Solano, D. A., B. E. Wohlberg, V. V. Vesselinov, and D. M. Tartakovsky (2015), Linear functional minimization for inverse modeling, *Water Resour. Res.*, *51*, 4516–4531, doi:10.1002/2014WR016179.

Received 2 AUG 2014
Accepted 23 MAY 2015
Accepted article online 1 JUN 2015
Published online 21 JUN 2015

Abstract We present a novel inverse modeling strategy to estimate spatially distributed parameters of nonlinear models. The maximum a posteriori (MAP) estimators of these parameters are based on a likelihood functional, which contains spatially discrete measurements of the system parameters and spatiotemporally discrete measurements of the transient system states. The piecewise continuity prior for the parameters is expressed via Total Variation (TV) regularization. The MAP estimator is computed by minimizing a nonquadratic objective equipped with the TV operator. We apply this inversion algorithm to estimate hydraulic conductivity of a synthetic confined aquifer from measurements of conductivity and hydraulic head. The synthetic conductivity field is composed of a low-conductivity heterogeneous intrusion into a high-conductivity heterogeneous medium. Our algorithm accurately reconstructs the location, orientation, and extent of the intrusion from the steady-state data only. Addition of transient measurements of hydraulic head improves the parameter estimation, accurately reconstructing the conductivity field in the vicinity of observation locations.

1. Introduction

Hydraulic and transport properties of subsurface environments, such as hydraulic conductivity, transmissivity, and dispersivity, are often highly heterogeneous. Ubiquitous spatial variability is a product of multiple geological, mechanical, and physicochemical processes that form subsurface environments [*Nœtinger et al*., 2005; *Carrera et al*., 2005]. Heterogeneity is observed over a large range of scales, from well-defined large geological structures with significantly different hydrogeological properties to smaller variations of these properties inside each structure. Accounting for multiple scales of heterogeneity in complex geologic environments is a prerequisite for construction of accurate and reliable predictive models [*Winter et al*., 2003; *de Marsily et al*., 2005].

Once a conceptual model of the subsurface processes has been constructed and a set of corresponding parameters has been identified, the next step is to estimate the values of these parameters via an appropriate inversion process that leverages both available field observations and prior knowledge and expectations, such as large-scale delineation of structures from geological analyses and models. Comprehensive reviews of alternative inversion strategies can be found in *Carrera et al*. [2005] and *Franssen et al*. [2009]. Most of these techniques are designed to handle a single heterogeneity scale [*Wohlberg et al*., 2012]: either single-facies heterogeneous systems [*Franssen et al*., 2009] or multiple homogeneous facies [*Iglesias and McLaughlin*, 2011; *Wohlberg et al*., 2006; *Tartakovsky and Wohlberg*, 2004]. A two-scale alternative [*Hu*, 2002; *Hu and Le Ravalec-Dupin*, 2004] combines a zonation approach for large-scale facies delineation with a stochastic description of within-facies variability. This strategy is strongly dependent on the selection of an adequate stochastic model.

We present an inversion methodology, which uses observations of both system parameters and transient system states to estimate the underlying parameter fields. No strong assumptions, such as prescribed zonation or stochastic models for small-scale heterogeneity, are introduced. Instead, we adopt a Bayesian framework, which incorporates the knowledge of spatial distribution of system parameters. Specifically, we chose a piecewise continuous representation of these parameters as their priors. The total variation (TV) regularization [*Rudin et al*., 1992] is used to define the prior with desired properties. A maximum a posteriori (MAP) estimator of the system parameters is obtained by solving the associated high-dimensional nonlinear

optimization problem. The linearized functional minimization algorithm [*Wohlberg et al.*, 2012] is used in this task. This algorithm splits the nonlinearities of the log-posterior into two parts, each of which is treated separately in an iterative scheme: the nonlinearity of the data fidelity terms associated with system states is treated using a linearization approach similar to the Levenberg-Marquardt method [*Bube and Langan*, 1997; *Bachmayr and Burger*, 2009]; and the nonlinearity associated with the TV regularization operator is dealt with using the alternating direction method of multipliers (ADMM) [*Boyd et al.*, 2010].

Our implementation of the TV regularization is different from those used for inverse modeling of spatially distributed parameters of elliptic equations [*Ascher and Haber*, 2004; *Chung et al.*, 2005; *Bachmayr and Burger*, 2009]. Furthermore, our methodology allows incorporation of transient measurements of various system states. This feature is essential for history matching, where information of the response of measured system states can be leveraged to improve the reconstruction of the model parameters. Our implementation uses measurements of a model parameter and a single state, but can be extended to include multiple states. *Lee and Kitanidis* [2013] have employed the TV regularization as a prior for Bayesian inverse modeling and structure identification in the context of hydrogeological applications. Their work employs a TV prior to sample the Bayesian posterior, and does not use transient measurements of system states. In contrast, we focus on efficient computation of the MAP estimator defined by the Bayesian posterior via an optimization algorithm.

Section 2 presents our inversion methodology in the context of saturated flow. The linearized functional minimization algorithm is described in section 3. The numerical schemes used both to solve the forward problem and to compute sensitivities are outlined in section 4. Computational cost of the overall inversion procedure is analyzed in section 5. In section 6, we apply our inversion methodology to a synthetic two-dimensional saturated flow problem with a highly heterogeneous, piecewise continuous conductivity field. Conductivity data are collected over a grid of discrete sampling locations, together with measurements of hydraulic head at a number of discrete locations. Conclusions and implications of our work are presented in section 7.

2. Problem Formulation

We consider an n-dimensional groundwater flow equation

$$\frac{\partial h}{\partial t} = \nabla \cdot (K \nabla h) + q(\mathbf{x}, t), \tag{1}$$

subject to appropriate boundary and initial conditions. Here $K(\mathbf{x})$ is the spatially varying saturated hydraulic conductivity (a system parameter), $h(\mathbf{x}, t)$ is the hydraulic head (a system state), and the source term $q(\mathbf{x}, t)$ represents, e.g., pumping wells. A spatiotemporal discretization of (1) replaces the continuous functions $K(\mathbf{x})$ and $h(\mathbf{x}, t)$ with their discrete counterparts arranged in vectors $\mathbf{k}$ and $\mathbf{h}$, respectively. We employ a method-of-lines (MOL) discretization, in which $K(\mathbf{x})$ is evaluated at N_k discrete grid points in a computational domain and $h(\mathbf{x}, t)$ is computed at N_h discrete grid points in space and N_t discrete points in time. As a result, vectors $\mathbf{k}$ and $\mathbf{h}$ have N_k and $N_h N_t$ components, respectively. Combined with the initial and boundary conditions, (1) defines a nonlinear map $\mathbf{h} = \mathbf{f}(\mathbf{k})$.

Suppose that hydraulic conductivity $K(\mathbf{x})$ is measured at $N_{\text{obs},k}$ grid points ($N_{\text{obs},k} \ll N_k$) of the numerical mesh used to discretize (1). Suppose further that hydraulic head $h(\mathbf{x}, t)$ is measured at $N_{\text{obs},h}$ grid points at each of the N_t discrete time points. The support volume of the conductivity measurements is considered representative of the cell size in the numerical model. These measurements are assembled into vectors $\hat{\mathbf{k}}$ and $\hat{\mathbf{h}}$ whose dimensions are $N_{\text{obs},k}$ and $N_{\text{obs},h} N_t$, respectively. Linear operators $\mathbf{M}_k$ and $\mathbf{M}_h$ provide spatiotemporal maps between $\mathbf{k}$ and $\hat{\mathbf{k}}$ and between $\mathbf{h}$ and $\hat{\mathbf{h}}$, such that $\mathbf{M}_k \mathbf{k} = \hat{\mathbf{k}}$ and $\mathbf{M}_h \mathbf{h} = \hat{\mathbf{h}}$. The measurement operator matrices $\mathbf{M}_k$ and $\mathbf{M}_h$ have dimensions $N_{\text{obs},k} \times N_k$ and $N_{\text{obs},h} N_t \times N_h N_t$, respectively. Errors in measurement of hydraulic conductivity and hydraulic head are assumed to be normal with covariance matrices $\mathbf{\Sigma}_k$ and $\mathbf{\Sigma}_h$, respectively.

Our goal is to compute estimators of $\mathbf{k}$ and $\mathbf{h}$ using the nonlinear map $\mathbf{h} = \mathbf{f}(\mathbf{k})$ and the available measurements $\hat{\mathbf{k}}$ and $\hat{\mathbf{h}}$. Since the measurement operator $\mathbf{M}_k$ is in general highly undetermined, it is not possible to

use direct inversion or to compute an ordinary least squares estimator of $\mathbf{k}$ from $\hat{\mathbf{k}}$. Instead, we pose an inversion problem by including the state variable data $\hat{\mathbf{h}}$, together with the available general knowledge of the properties of $K(\mathbf{x})$ and $h(\mathbf{x},t)$.

Following *Wohlberg et al*. [2012], we express the inversion problem in a variational form as the minimization of a (negative) likelihood functional. This likelihood is composed of data fidelity terms, which penalize the difference between model estimates and measurements, and regularization terms, which express the prior knowledge of the properties of the (discretized) fields $\mathbf{k}$ and $\mathbf{h}$. Enforcing the constraint $\mathbf{h}=\mathbf{f}(\mathbf{k})$, that gives rise to an optimization problem over $\mathbf{k}$,

$$\mathbf{k}_{\mathrm{MAP}}=\arg\min_{\mathbf{k}}\frac{1}{2}\left(\mathbf{M}_k\mathbf{k}-\hat{\mathbf{k}}\right)^{\top}\boldsymbol{\Sigma}_k\left(\mathbf{M}_k\mathbf{k}-\hat{\mathbf{k}}\right)+\frac{1}{2}\left(\mathbf{M}_h\mathbf{f}(\mathbf{k})-\hat{\mathbf{h}}\right)^{\top}\boldsymbol{\Sigma}_h\left(\mathbf{M}_h\mathbf{f}(\mathbf{k})-\hat{\mathbf{h}}\right)+\gamma R_k(\mathbf{k})+\frac{\delta}{2}R_h(\mathbf{f}(\mathbf{k})), \quad (2)$$

where $R_k(\mathbf{k})$ and $R_h(\mathbf{h}=\mathbf{f}(\mathbf{k}))$ are the regularization terms for $\mathbf{k}$ and $\mathbf{h}$, respectively. The estimators $\mathbf{k}_{\mathrm{MAP}}$ and $\mathbf{h}_{\mathrm{MAP}}=\mathbf{f}(\mathbf{k}_{\mathrm{MAP}})$ are known as maximum a posteriori (MAP) estimators of $\mathbf{k}$ and $\mathbf{h}$.

Without loss of generality, we assume that observation errors are independent and that $(\Sigma_k)_{ii}=\alpha^{-1}$ and $(\Sigma_h)_{ii}=\beta^{-1}$, where $\alpha,\beta>0$ are two constants. Then the optimization problem (2) reduces to

$$\mathbf{k}_{\mathrm{MAP}}=\arg\min_{\mathbf{k}}\frac{\alpha}{2}||\mathbf{M}_k\mathbf{k}-\hat{\mathbf{k}}||_2^2+\frac{\beta}{2}||\mathbf{M}_h\mathbf{f}(\mathbf{k})-\hat{\mathbf{h}}||_2^2+\gamma R_k(\mathbf{k})+\frac{\delta}{2}R_h(\mathbf{f}(\mathbf{k})). \quad (3)$$

From the Bayesian perspective, the data fidelity terms provide the likelihood of observations given a certain configuration of $\mathbf{k}$, and the regularization terms provide the priors on $\mathbf{k}$ and $\mathbf{h}$. The inversion coefficients α, β, γ, and δ determine the extent to which low probability is assigned to a certain $\mathbf{k}$ configuration, i.e., the relative contributions of each term in (3). As indicated before, the coefficients α and β are proportional to the inverse of the variance of the error of the observations of $\mathbf{k}$ and $\mathbf{h}$, respectively. Similarly, γ and δ correspond to the degree of confidence in the prior hypothesis captured by the respective regularization terms. Since we are not interested in the values of the likelihood, only relative values of the inversion coefficients are relevant to the minimization process. Therefore, without loss of generality, we set $\alpha=1$.

In subsurface environments consisting of multiple facies (see section 1), the hydraulic conductivity field $K(\mathbf{x})$ is piecewise continuous. To reflect this knowledge, we select the regularization for $\mathbf{k}$ to be given by the ℓ_1 norm of the gradient of the conductivity field ∇K. This regularization is often referred to as total variation (TV) [*Rudin et al*., 1992]. In n-dimensional domains, it penalizes general heterogeneous fields and promotes fields with low small-scale variation inside facies and large jumps across the $(n-1)$-dimensional surfaces separating the facies. Since $\mathbf{h}$ is continuous and defined by the map $\mathbf{h}=\mathbf{f}(\mathbf{k})$, we impose no additional regularity requirement on $\mathbf{h}$. Instead, we add an additional regularization term for $\mathbf{k}$, which penalizes the ℓ_2 norm of ∇K. This increases stability during the minimization of (3). To avoid interference with the regularization properties of the TV norm, the ℓ_2 regularization term is multiplied by a coefficient δ, which satisfies the inequality $\delta \ll \gamma$.

Two sources contribute to high nonlinearity of the objective functional (3): the data fidelity penalty on the deviation of the modeled head from observations, $||\mathbf{M}_h\mathbf{f}(\mathbf{k})-\hat{\mathbf{h}}||_2^2$, and the TV norm. The latter also renders the optimization functional nondifferentiable with respect to $\mathbf{k}$. Without the TV norm, the problem is equivalent to a nonlinear least squares estimation, which can be solved by a variety of standard techniques (e.g., Levenberg-Marquardt). On the other hand, the problem without the nonquadratic penalty is equivalent to linear minimization equipped with the TV norm, for which many efficient numerical approaches have been proposed [*Rodríguez and Wohlberg*, 2009; *Goldstein and Osher*, 2009; *Boyd et al*., 2010]. We pursue a hybrid approach [*Wohlberg et al*., 2012] that combines the strengths of the methods designed for dealing with each type of nonlinearity. This strategy separates the two sources of nonlinearity by first linearizing the map $\mathbf{h}=\mathbf{f}(\mathbf{k})$ and then minimizing a sequence of the linearized models equipped with the TV norm.

For a current estimate $\mathbf{k}^{(c)}$ of the minimizer, we replace $\mathbf{f}^{(c)}$ with the affine model [*Bube and Langan*, 1997; *Bachmayr and Burger*, 2009]

$$\mathbf{f}(\mathbf{k}) \approx \mathbf{f}(\mathbf{k}^{(c)}) + \mathbf{J}_{\mathbf{f}}(\mathbf{k}^{(c)})(\mathbf{k}-\mathbf{k}^{(c)}), \tag{4}$$

where $\mathbf{J}_{\mathbf{f}}(\mathbf{k}^{(c)}) = \partial\mathbf{f}/\partial\mathbf{k}^{(c)}$ is the Jacobian of $\mathbf{f}$ evaluated at $\mathbf{k}^{(c)}$. Next, we approximate the head data fidelity term with

$$||\mathbf{M}_h\mathbf{f}(\mathbf{k}) - \hat{\mathbf{h}}||_2^2 \approx ||\mathbf{M}_h\mathbf{J}_{\mathbf{f}}(\mathbf{k}^{(c)})\mathbf{k} + \mathbf{M}_h\{\mathbf{f}(\mathbf{k}^{(c)}) - \mathbf{J}_{\mathbf{f}}(\mathbf{k}^{(c)})\mathbf{k}^{(c)}\} - \hat{\mathbf{h}}||_2^2. \tag{5}$$

The modified minimization problem associated with the iterate $\mathbf{k}^{(c)}$ is a quadratic form, except for the TV norm. Any of the previously mentioned TV optimization schemes can be used to solve this optimization problem. A solution of this linearized subproblem yields a new iterate $\mathbf{k}^{(c+1)}$, for which a new linearized minimization problem is solved, and so forth. This linearized functional minimization (LFM) algorithm is described below.

3. Linearized Functional Minimization

We use the LFM algorithm [*Wohlberg et al.*, 2012] to solve (3) together with a TV regularization term. In broad terms, this algorithm is a combination of the Levenberg trust-region algorithm, which is employed to propose linearized minimization subproblems, and the Alternating Direction Method of Multipliers (ADMM) approach [*Boyd et al.*, 2010], which is used to solve these subproblems.

Consider a generalized problem with a parameter vector $\mathbf{u} \in \mathbb{R}^N$ and a state vector $\mathbf{v} \in \mathbb{R}^M$, stemming from a suitable discretization of a partial-differential equation (PDE) on a two-dimensional domain Ω and time interval $(0, T]$. The discretized PDE provides a map $\mathbf{v}=\mathbf{f}(\mathbf{u})$. Let $\mathbf{P}$ and $\mathbf{Q}$ denote measurement operators such that $\mathbf{Pu}=\mathbf{s}$ and $\mathbf{Qv}=\mathbf{t}$, with $\mathbf{s} \in \mathbb{R}^{N'}$ and $\mathbf{t}=\mathbb{R}^{M'}$, and $M' \ll M$ and $N' \ll N$. The cost functional (3) with a TV regularization term takes the form

$$\arg\min_{\mathbf{u}} \frac{\alpha}{2}||\mathbf{Pu}-\mathbf{s}||_2^2 + \frac{\beta}{2}||\mathbf{Qf}(\mathbf{u})-\mathbf{t}||_2^2 + \gamma||\mathcal{D}(\mathbf{u})||_1 + \frac{\delta}{2}||\mathcal{D}(\mathbf{u})||_2^2, \tag{6}$$

where $\mathcal{D}(\cdot) = \sqrt{[\mathbf{D}_x(\cdot)]^2 + [\mathbf{D}_y(\cdot)]^2}$ is a discretized gradient magnitude operator, with the operations $[\cdot]^2$ and $\sqrt{\cdot}$ understood as element-wise, and $\mathbf{D}_x$ and $\mathbf{D}_y$ are linear discrete differentiation operators in the x and y directions, respectively.

We use an iterative trust region approach to minimize (6). Starting from an iterate $\mathbf{u}^{(c)}$, we define the next iterate as $\mathbf{u}^{(+)}=\mathbf{u}^{(c)}+\mathbf{w}$, where $\mathbf{w}$ is chosen so as to minimize a suitably chosen linearization of (6). For a $\mathbf{w}$ with small squared L^2-norm, $||\mathbf{w}||_2^2$, we employ an affine model

$$\mathbf{v} \approx \mathbf{v}^{(c)} + \mathbf{Aw},$$

where $\mathbf{v}^{(c)}=\mathbf{f}(\mathbf{u}^{(c)})$, and $\mathbf{A}$ is the Jacobian matrix of $\mathbf{f}(\mathbf{u}^{(c)})$, whose elements are $A_{pq}=\partial f_p(\mathbf{u}^{(c)})/\partial u_q$. The step $\mathbf{w}$ is chosen as the minimizer of

$$\arg\min_{\mathbf{w}} \frac{\alpha^*}{2}||\mathbf{P}(\mathbf{u}^{(c)}+\mathbf{w})-\mathbf{s}||_2^2 + \frac{\beta^*}{2}||\mathbf{Q}(\mathbf{v}^{(c)}+\mathbf{Aw})-\mathbf{t}||_2^2 + \gamma^*||\mathcal{D}(\mathbf{u}^{(c)}+\mathbf{w})||_1 + \frac{\delta^*}{2}||\mathcal{D}(\mathbf{u}^{(c)}+\mathbf{w})||_2^2 + \frac{1}{2}||\mathbf{w}||_2^2, \tag{7}$$

where the penalty on $||\mathbf{w}||_2^2$ is added to provide regularity (convexity) of the solution. The inversion coefficients of the linearized problem, α^*, β^*, γ^*, and δ^*, are chosen initially to coincide with their counterparts in (6). They can also be allowed to vary in conjunction in order to control the trust region size throughout the minimization process.

The ADMM approach [*Boyd et al.*, 2010] is used to minimize the linearized functional (7). This allows us to split the regularization terms from the deviation penalties. Introducing auxiliary variables $\mathbf{d}_x$ and $\mathbf{d}_y$, we recast (7) as

$$\arg\min_{\mathbf{w}} \frac{\alpha^*}{2}||\mathbf{P}(\mathbf{u}^{(c)}+\mathbf{w})-\mathbf{s}||_2^2 + \frac{\beta^*}{2}||\mathbf{Q}(\mathbf{v}^{(c)}+\mathbf{Aw})-\mathbf{t}||_2^2 + \gamma^*\left\|\sqrt{\mathbf{d}_x^2+\mathbf{d}_y^2}\right\|_1 + \frac{\delta^*}{2}\left\|\sqrt{\mathbf{d}_x^2+\mathbf{d}_y^2}\right\|_2^2 + \frac{1}{2}||\mathbf{w}||_2^2, \tag{8}$$

subject to $\mathbf{d}_x=\mathbf{D}_x(\mathbf{u}^{(c)}+\mathbf{w})$ and $\mathbf{d}_y=\mathbf{D}_y(\mathbf{u}^{(c)}+\mathbf{w})$. This constrained problem is solved iteratively via a splitting scheme

$$\begin{aligned}\mathbf{w}^{(k+1)} \equiv \arg\min_{\mathbf{w}} &\frac{\lambda}{2}||\mathbf{D}_x\mathbf{w}-(\mathbf{d}_x^{(k)}-\mathbf{D}_x\mathbf{u}^{(c)}-\mathbf{b}_x^{(k)})||_2^2+\frac{\lambda}{2}||\mathbf{D}_y\mathbf{w}-(\mathbf{d}_y^{(k)}-\mathbf{D}_y\mathbf{u}^{(c)}-\mathbf{b}_y^{(k)})||_2^2\\ &+\frac{1}{2}||\mathbf{w}||_2^2+\frac{\alpha^*}{2}||\mathbf{P}\mathbf{w}-(\mathbf{s}-\mathbf{P}\mathbf{u}^{(c)})||_2^2+\frac{\beta^*}{2}||\mathbf{Q}\mathbf{A}\mathbf{w}-(\mathbf{t}-\mathbf{Q}\mathbf{v}^{(c)})||_2^2,\end{aligned} \tag{9}$$

and

$$\begin{aligned}(\mathbf{d}_x^{(k+1)},\mathbf{d}_y^{(k+1)}) \equiv \arg\min_{\mathbf{d}_x,\mathbf{d}_y} &\gamma^*\|\sqrt{\mathbf{d}_x^2+\mathbf{d}_y^2}\|_1+\frac{\delta^*}{2}\|\sqrt{\mathbf{d}_x^2+\mathbf{d}_y^2}\|_2^2+\frac{\lambda}{2}||\mathbf{D}_x\mathbf{w}^{(k+1)}-(\mathbf{d}_x-\mathbf{D}_x\mathbf{u}^{(c)}-\mathbf{b}_x^{(k)})||_2^2\\ &+\frac{\lambda}{2}||\mathbf{D}_y\mathbf{w}^{(k+1)}-(\mathbf{d}_y-\mathbf{D}_y\mathbf{u}^{(c)}-\mathbf{b}_y^{(k)})||_2^2,\end{aligned} \tag{10}$$

where $\lambda>0$. The dual variables $\mathbf{b}_x$ and $\mathbf{b}_y$ are updated as

$$\mathbf{b}_x^{(k+1)}=\mathbf{b}_x^{(k)}+\mathbf{D}_x(\mathbf{u}^{(c)}+\mathbf{w}^{(k+1)})-\mathbf{d}_x^{(k+1)}, \tag{11}$$

$$\mathbf{b}_y^{(k+1)}=\mathbf{b}_y^{(k)}+\mathbf{D}_y(\mathbf{u}^{(c)}+\mathbf{w}^{(k+1)})-\mathbf{d}_y^{(k+1)}. \tag{12}$$

Initial values for the iteration scheme are set to $\mathbf{b}_x^{(0)}=0$, $\mathbf{b}_y^{(0)}=0$, $\mathbf{d}_x^{(0)}=\mathbf{D}_x\mathbf{u}^{(c)}$, and $\mathbf{d}_y^{(0)}=\mathbf{D}_y\mathbf{u}^{(c)}$.

A solution to the linear minimization problem (9) is obtained by differentiating and setting equal to 0 the argument of the right-hand side of (9), and then solving the resulting $N\times N$ linear algebraic system

$$(\mathbf{\Lambda}+\beta^*\mathbf{A}^\top\mathbf{Q}^\top\mathbf{Q}\mathbf{A})\mathbf{w}^{(k+1)}=\mathbf{r}^{(k)}, \tag{13}$$

where

$$\mathbf{\Lambda}=\mathbf{I}_N+\alpha^*\mathbf{P}^\top\mathbf{P}+\lambda\mathbf{D}_x^\top\mathbf{D}_x+\lambda\mathbf{D}_y^\top\mathbf{D}_y, \tag{14}$$

$\mathbf{I}_N$ is the $N\times N$ identity matrix, and

$$\mathbf{r}^{(k)}=\alpha^*\mathbf{P}^\top(\mathbf{s}-\mathbf{P}\mathbf{u}^{(c)})+\beta^*\mathbf{A}^\top\mathbf{Q}^\top(\mathbf{t}-\mathbf{Q}\mathbf{v}^{(c)})+\lambda\mathbf{D}_x^\top(\mathbf{d}_x^{(k)}-\mathbf{D}_x\mathbf{u}^{(c)}-\mathbf{b}_x^{(k)})+\lambda\mathbf{D}_y^\top(\mathbf{d}_y^{(k)}-\mathbf{D}_y\mathbf{u}^{(c)}-\mathbf{b}_y^{(k)}). \tag{15}$$

The problem for $\mathbf{d}_x^{(k+1)}$ and $\mathbf{d}_y^{(k+1)}$ is solved component-wise to yield

$$d_{xi}^{(k+1)}=h_{xi}\sigma(h_{xi},h_{yi},\gamma^*/\lambda,\delta^*/\lambda), \tag{16}$$

and

$$d_{yi}^{(k+1)}=h_{yi}\sigma(h_{xi},h_{yi},\gamma^*/\lambda,\delta^*/\lambda), \tag{17}$$

where $\mathbf{h}_x=\mathbf{D}_x(\mathbf{u}^{(c)}+\mathbf{w}^{(k+1)})+\mathbf{b}_x^{(k)}$, $\mathbf{h}_y=\mathbf{D}_y(\mathbf{u}^{(c)}+\mathbf{w}^{(k+1)})+\mathbf{b}_y^{(k)}$, and σ is the shrinkage function [*Goldstein and Osher*, 2009] defined as

$$\sigma(a,b,\xi,\eta)=\frac{\max(0,\sqrt{a^2+b^2}-\xi)}{(1+\eta)\sqrt{a^2+b^2}}. \tag{18}$$

The splitting scheme can be generalized to three-dimensional domains, in which case one introduces an additional variable $\mathbf{d}_z$ subject to $\mathbf{d}_z=\mathbf{D}_z(\mathbf{u}^{(c)}+\mathbf{w})$ into the splitting. The $(\mathbf{d}_x,\mathbf{d}_y,\mathbf{d}_z)$ subproblem is solved via the ansatz $d_{xi}=h_{xi}\sigma$, $d_{yi}=h_{yi}\sigma$, and $d_{zi}=h_{zi}\sigma$, which leads to a shrinkage formula similar to (18) except with $\sqrt{a^2+b^2+c^2}$ instead of $\sqrt{a^2+b^2}$.

The iteration procedure (9)–(12) is repeated until the relative change in the functional (7) between iterations falls below a certain tolerance. Completing this iteration procedure produces a candidate $\mathbf{w}$, which is accepted or rejected as follows. (i) The candidate is accepted if it produces a reduction of the functional (6) and rejected otherwise. (ii) If the candidate is rejected, the coefficients α^*, β^*, γ^*, and δ^* are reduced by a factor of 2 (thus reducing the trust region) and a new candidate is computed. This process is repeated until

a satisfactory candidate is found. (iii) If the first candidate **w** for a certain starting point is satisfactory, the coefficients α^*, β^*, γ^*, and δ^* are increased by a factor of 2 (thus expanding the trust region).

The minimization of (6) is carried until the relative change in the functional falls below a certain tolerance. The vector $\mathbf{u}^{(+)}=\mathbf{u}^{(c)}+\mathbf{w}$ of the last iteration is taken as the estimator of the parameter vector.

4. Sensitivity Analysis

The methodology described in the previous section requires computing the state and the Jacobian associated with each iterate of the model parameter vector **u**. We employ a method-of-lines (MOL) strategy to solve the flow equation (1). An MOL discretization of (1) produces a system of N_h linear ODEs

$$\left[\mathbf{V}\frac{\mathrm{d}}{\mathrm{d}t}+\mathbf{F}(\mathbf{k})\right]\tilde{\mathbf{h}}(t)=\mathbf{g}(\mathbf{k})+\mathbf{q}, \tag{19}$$

where $\tilde{\mathbf{h}}(t)$ is the hydraulic head vector at time t, **V** and **F** are the corresponding "capacitance" and "conductivity" matrices of the discretization scheme, **g** is the forcing vector associated with the boundary conditions, and **q** is the forcing vector representing the pumping rates.

The set of ODEs (19) is integrated in time using an appropriate time stepping scheme, and the state vector is stored for the N_t discrete observation times $t_1, t_2, \ldots t_{N_t}$. The full state vector is assembled as $\mathbf{h}=[\tilde{\mathbf{h}}(t_1)^\top, \tilde{\mathbf{h}}(t_2)^\top, \ldots \tilde{\mathbf{h}}(t_{N_t})^\top]^\top$.

For stationary analysis, (19) reduces to the linear system of algebraic equations

$$\mathbf{F}(\mathbf{k})\mathbf{h}=\mathbf{g}(\mathbf{k})+\mathbf{q}, \tag{20}$$

where **h** is the stationary hydraulic head vector, and **q** is the stationary forcing vector representing the aquifer pumping rates.

4.1. Forward Sensitivity Analysis

Assuming that the aquifer pumping rates vector **q** is independent of **k**, and using the chain rule with respect to the ith element of **k** in (19), we obtain a set of linear ODEs governing the evolution of the ith column of the sensitivity matrix $\mathbf{J}_{,i}=\partial\tilde{\mathbf{h}}/\partial k_i$:

$$\left[\mathbf{V}\frac{\mathrm{d}}{\mathrm{d}t}+\mathbf{F}(\mathbf{k})\right]\mathbf{J}_{,i}=\frac{\partial\mathbf{g}(\mathbf{k})}{\partial k_i}-\frac{\partial\mathbf{F}(\mathbf{k})}{\partial k_i}\tilde{\mathbf{h}}. \tag{21}$$

Let **G** be an $N_h\times N_k$ matrix whose columns are given by the right-hand side of (21),

$$\mathbf{G}_{,i}(\tilde{\mathbf{h}})=\frac{\partial\mathbf{g}(\mathbf{k})}{\partial k_i}-\frac{\partial\mathbf{F}(\mathbf{k})}{\partial k_i}\tilde{\mathbf{h}}. \tag{22}$$

Then, (21) is rewritten as

$$\left[\mathbf{V}\frac{\mathrm{d}}{\mathrm{d}t}+\mathbf{F}(\mathbf{k})\right]\mathbf{J}=\mathbf{G}(\tilde{\mathbf{h}}). \tag{23}$$

Computation of the sensitivity matrix requires the integration in time of N_k sets of N_h ODEs. For each forward simulation, the sensitivities at the N_t observation times are stored, and the Jacobian **A** of the functional $\mathbf{h}=\mathbf{f}(\mathbf{k})$ is assembled as $\mathbf{A}=[\mathbf{J}(t_1)^\top, \mathbf{J}(t_2)^\top, \ldots \mathbf{J}(t_{N_t})^\top]^\top$.

The sets of ODEs (19) and (23) are similar, differing only in their arguments and right-hand sides. Therefore, computation of the state vector and Jacobian can be understood as the simultaneous time stepping of $N_h(N_k+1)$ linear ODEs. Given that one of the leading costs of the time stepping of linear systems of ODEs using implicit schemes is the factorization of the time stepping matrix, time integration of state and sensitivities together leads to significant computational savings.

For stationary analysis, the ith column of the sensitivity matrix is given by the linear system of equations

$$\mathbf{F}(\mathbf{k})\mathbf{J}=\mathbf{G}(\mathbf{h}), \tag{24}$$

where **h** is the solution to (20).

4.2. Adjoint Sensitivity Analysis

A solution to the set of ODEs (20) and the linear system of equations (24) provides the complete sensitivity information of the system codified in the Jacobian matrix $\mathbf{A}$ of size $N_h N_t \times N_k$, i.e., the sensitivity of all hydraulic head degrees of freedom to the spatial configuration of hydraulic conductivity. On the other hand, the linearized functional minimization algorithm of section 3 requires the restricted Jacobian $\mathbf{M}_h\mathbf{A}$ of smaller dimension $N_{\text{obs},h} N_t \times N_k$, i.e., only the sensitivity of the hydraulic head at the given spatiotemporal observation location.

The various rows of the restricted Jacobian $\mathbf{M}_h\mathbf{A}$ can be computed directly via adjoint sensitivity analysis without having to solve the forward problems (23) and (24) for the full Jacobian. This alternative approach can lead to a significant reduction in computational cost per iteration step of the LFM algorithm if the number of observation locations of head is much smaller than the number of degrees of freedom of the discrete conductivity field.

In the stationary case, $\mathbf{M}_h$ is of dimension $N_{\text{obs},h} \times N_h$, and $\mathbf{A}=\mathbf{J}$. Let $\mathcal{I}_h=[1, N_h]$, and $\mathcal{I}_{\text{obs},h} \subset \mathcal{I}_h$ be the subset of indices corresponding to observation locations. We denote these indices by l_i, $i \in [1, N_{\text{obs},h}]$. Let $\mathbf{e}_{l_i}$ be the $N_h \times 1$ vector with components $(e_{l_i})_j=1$ if $j=l_i$ and zero otherwise. By construction, $\mathbf{e}_{l_i}^\top$ corresponds to the ith row of $\mathbf{M}_h$. Left-multiplying (24) with $\mathbf{F}^{-1}$ and $\mathbf{e}_{l_i}^\top$, we obtain the following expression for the l_i, m component of the Jacobian,

$$\frac{\partial h_{l_i}}{\partial k_m}=J_{l_i,m}=\mathbf{e}_{l_i}^\top \mathbf{F}(\mathbf{k})^{-1}\mathbf{G}(\mathbf{h})_{,m}. \tag{25}$$

Let us introduce the set of adjoint vectors $\boldsymbol{\mu}_{l_i}$ ($l_i \in \mathcal{I}_{\text{obs},h}$), defined as $\boldsymbol{\mu}_{l_i}^\top=\mathbf{e}_{l_i}\mathbf{F}^{-1}$. We assemble the various adjoint vectors into the $N_{\text{obs},h} \times N_h$ matrix $\boldsymbol{\mu}=[\boldsymbol{\mu}_{l_1}, \ldots, \boldsymbol{\mu}_{l_{N_{\text{obs},h}}}]$ in order to obtain the adjoint linear system of equations

$$\boldsymbol{\mu}^\top \mathbf{F}(\mathbf{k})=\mathbf{M}_h. \tag{26}$$

Combining (26) and (25) we obtain

$$\mathbf{M}_h\mathbf{J}=\boldsymbol{\mu}^\top \mathbf{G}(\mathbf{h}). \tag{27}$$

Compared to solving (24) directly, computing $\boldsymbol{\mu}$ via (26) requires the solution of $N_{\text{obs},h}$ linear systems of equations of dimension $N_h \times N_h$, as opposed to N_k. This leads to a significant reduction in computational cost if $N_{\text{obs},h} \ll N_k$ and N_h is large.

In the transient case, we have $\mathbf{M}_h=\mathbf{I}_{N_t} \otimes \mathbf{M}_{\tilde{h}}$, where $\mathbf{I}_{N_t}$ is the $N_t \times N_t$ identity matrix, and $\mathbf{M}_{\tilde{h}}$ is the $N_{\text{obs},h} \times N_h$ spatial observation operator. Similar to the stationary case, it is possible to compute the product $\mathbf{M}_{\tilde{h}}\mathbf{J}(T)$ directly without having to explicitly compute $\mathbf{J}(T)$ by solving an adjoint problem of smaller dimension [*Cao et al.*, 2003].

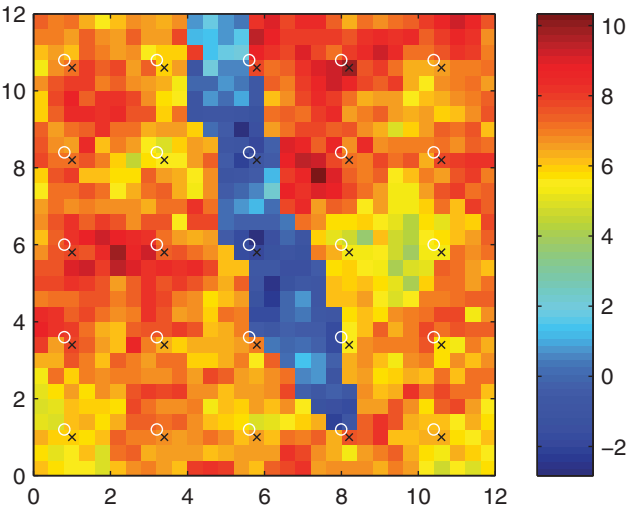

Figure 1. Synthetic log hydraulic conductivity field for test case 1. Hydraulic conductivity observation locations are indicated with crosses. Hydraulic head observation locations are indicated with circles.

The $N_{\text{obs},h} \times N_h$ adjoint variable $\boldsymbol{\mu}$ is defined as a solution of the ODE set adjoint to (19), together with a terminal condition at $t=T$, i.e.,

$$\frac{\mathrm{d}\boldsymbol{\mu}^\top(t)}{\mathrm{d}t}\mathbf{V}-\boldsymbol{\mu}^\top(t)\mathbf{F}(\mathbf{k})=0, \quad \boldsymbol{\mu}^\top(T)\mathbf{V}=\mathbf{M}_{\tilde{h}}. \tag{28}$$

The restricted Jacobian $\mathbf{M}_{\tilde{h}}\mathbf{J}(T)$ is given by

$$\mathbf{M}_{\tilde{h}}\mathbf{J}(T)=\boldsymbol{\mu}^\top(0)\mathbf{V}\mathbf{J}(0)+\int_0^T \boldsymbol{\mu}^\top(t)\mathbf{G}(\tilde{\mathbf{h}}(t))\,\mathrm{d}t, \tag{29}$$

where $\mathbf{J}(0)$ is the sensitivity of the initial condition $\tilde{\mathbf{h}}(0)$ to $\mathbf{k}$.

The time stepping of the adjoint ODEs (28) requires the solution of $N_{\text{obs},h} N_t$ linear

Table 1. Properties of Point Loadings of the Test Cases

Scenario	Pumping Well Location	Pumping Rate [$L\ T^{-1}$]
1	No pumping	
2	(6.0, 6.0)	5×10^{0}
3	(2.8, 6.0)	5×10^{3}

algebraic systems of size $N_h \times N_h$ per time step, compared to N_k systems for the forward ODEs (23). This leads to a significant reduction in computational cost if $N_{\text{obs},h} N_t \ll N_k$ and N_h large.

Computation of the restricted sensitivity matrix $\mathbf{M}_{\tilde{h}}\mathbf{J}(T)$ via (29) requires integrating the ODEs (19) forward in time from $t = 0$ to $t = T$, integrating the adjoint ODEs (28) backward in time from $t = T$ to $t = 0$, and storing both $\tilde{\mathbf{h}}(t)$ and $\boldsymbol{\mu}(t)$ at certain time steps in order to evaluate the quadrature in the right-hand side of (29). If the total number of stored time steps is large, this may result in expensive memory requirements. As a solution, one can employ the so-called checkpointing approach [*Serban and Hindmarsh*, 2005], which reduces the memory requirements by introducing at most one additional forward integration of the state.

5. Computational Cost

The feasibility of this inversion strategy depends on one's ability to efficiently solve both the forward problems and the iterative optimization subproblems. The most expensive operations of this method are (i) integration of the forward ODEs (19), (ii) integration of either forward (23) or backward (29) sensitivity problems, and (iii) solution of the linear algebraic system (13). Below we discuss these operations and their computational complexity as implemented in order to ensure adequate scaling with the number of observations and the number of degrees of freedom.

5.1. Outer Iteration Scaling

Time stepping of the ODEs using implicit methods requires solving the various linear algebraic systems of size $N_h \times N_h$ per time step. We therefore use high-order, variable-step ODE integrators for their solution. The use of high-order methods with variable time steps allows for accurate time stepping with long time steps, reducing the total amount of time steps and the overall cost. Several robust high-order implicit schemes are known to perform well on sets of ODEs stemming from MOL discretizations. These include the CVODES package [*Serban and Hindmarsh*, 2005], which uses high-order backward differences formulae (BDF), and Implicit Runge-Kutta (IRK) methods [*Hairer and Wanner*, 1999]. In the simulations reported below we used the three-stage, order 5 Radau IIA IRK formula, which allows for time step selection via a posteriori error control.

Following the discussion in section 4.1, the computational cost of forward integration of the state and Jacobian ODEs (19) and (23) is

$$\mathcal{C}_{\text{fwd}} = O(N_{\Delta t}(N_k + 1)N_h^{\eta}), \tag{30}$$

where $N_{\Delta t}$ is the number of time steps, and η is a scaling coefficient specific to the linear solver employed. For 2-D problems, we have $\eta \approx 1$ for both optimal iterative solvers and direct multifrontal triangular solvers for sparse matrices. The sparse solver also requires an additional $O(N_h^{3/2})$ work for factoring the time stepping matrix, which is done once for each outer iteration of the LFM algorithm.

For backward sensitivity analysis (section 4.2), i.e., forward integration of the state ODE (19) and backward integration of the adjoint ODE (29), the computational cost is

$$\mathcal{C}_{\text{bwd}} = O(N_{\Delta t}(N_{\text{obs},h} N_t + 2)N_h^{\eta}). \tag{31}$$

Comparison of (30) and (31) reveals that the backward sensitivity approach is preferable for computing the Jacobian matrix when the number of observations of hydraulic head is smaller than the number of degrees of freedom of the conductivity field, i.e., when $N_{\text{obs},h} N_t \ll N_k$. This condition generally holds for the inversion applications of interest.

Table 2. Inversion Parameter Sets for Test Case 1

Set	α	β	γ	δ
a	1.0	1.0	1.0×10^{-6}	2.0×10^{-8}
b	1.0	0.5	1.0×10^{-6}	2.0×10^{-8}
c	1.0	1.0	1.0×10^{-4}	2.0×10^{-8}
d	1.0	0.5	1.0×10^{-4}	2.0×10^{-8}

Our implementation allows one to control both the error of the state and the adjoint variable (or the Jacobian itself, for forward sensitivity analysis). Our experience shows that controlling

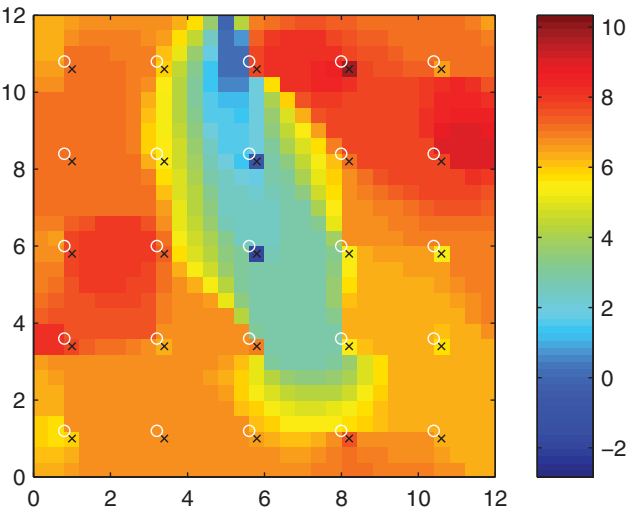

Figure 2. MAP estimator of hydraulic conductivity field for test case 1, Scenario 1, computed using the (a) inversion parameter set in Table 2.

for the Jacobian error tends to produce smaller steps than controlling solely for the state error, without any appreciable impact in the inversion procedure. Hence, we use only state error control. The accuracy of the Jacobian's approximation affects chiefly the computation of $\mathbf{w}^{(k+1)}$ in (9) and (13). If the approximation of $\mathbf{A}$ were to be insufficiently accurate to negatively affect the computation of $\mathbf{w}$, the candidate would still be rejected if it does not decrease the objective functional (6). Nevertheless, our experience shows that this is not the case: controlling only for the error of the state produces adequate approximations of $\mathbf{A}$.

5.2. Inner Iteration Scaling

The linear system (13) is of size $N_k \times N_k$. The Jacobian matrix $\mathbf{A}$ is a dense matrix, and so is (to a lesser extent) the product $\mathbf{A}^\top\mathbf{Q}^\top\mathbf{QA}$, which makes the solution of the system (13) expensive. Fortunately, this operation can be recast as a sequence of less expensive operations. By repeated applications of the matrix inversion lemma we rewrite (13) as

$$\mathbf{\Lambda c}_1 = \mathbf{r}^{(k)}, \tag{32}$$

$$\mathbf{\Lambda C}_1 = \mathbf{A}^\top\mathbf{Q}^\top, \tag{33}$$

$$((\beta^*)^{-1}\mathbf{I}_{N_{\mathrm{obs},h}N_t} + \mathbf{QAC}_1)\mathbf{c}_2 = \mathbf{QAc}_1, \tag{34}$$

$$\mathbf{\Lambda c}_3 = \mathbf{A}^\top\mathbf{Q}^\top\mathbf{c}_2, \tag{35}$$

$$\mathbf{w}^{(k+1)} = \mathbf{c}_1 - \mathbf{c}_3, \tag{36}$$

where $\mathbf{I}_{N_{\mathrm{obs},h}N_t}$ is the $N_{\mathrm{obs},h}N_t \times N_{\mathrm{obs},h}N_t$ identity matrix. This procedure requires the solution of $N_{\mathrm{obs},h}N_t + 2$ sparse linear algebraic systems of size $N_k \times N_k$ of the form $\mathbf{\Lambda x} = \mathbf{b}$, plus a smaller, dense system of size $N_{\mathrm{obs},h}N_t \times N_{\mathrm{obs},h}N_t$, which results in a more favorable scaling of computational cost when $N_{\mathrm{obs},h}N_t \ll N_k$.

The sparsity pattern of $\mathbf{\Lambda}$ is that of the discrete Laplacian operator, i.e., pentadiagonal and heptadiagonal for structured 2-D and 3-D hydraulic conductivity grids, respectively, allowing for the efficient solution of the systems (32), (33), and (35). When using direct methods, the factorization of $\mathbf{\Lambda}$ needs to be recomputed only when there is a change in the trust region and the coefficient α^* changes.

6. Application

The inversion strategy is tested with a synthetic hydraulic conductivity field in the square domain $[0, 12] \times [0, 12]$ $[L]$ with a resolution of 30 conductivity cells per spatial direction. The reference log conductivity field (Figure 1) is made of two heterogeneous facies with spatial averages differing by several orders of magnitude. Measurements of hydraulic conductivity are taken at $N_{\mathrm{obs},k} = 25$ cells. Note that there are only two head and conductivity observations in the low-conductivity region.

Table 3. Penalty Results for Test Case 1, Pumping Scenario 2, Computed Using the Inversion Parameters of Table 2

Set	$\|\|\mathbf{Pu}-\mathbf{s}\|\|_2^2/2$	$\|\|\mathbf{Qv}-\mathbf{t}\|\|_2^2/2$	TV $\|\|\mathcal{D}(\mathbf{u})\|\|_1$	$\|\|\mathcal{D}(\mathbf{u})\|\|_2^2$
a	9.69×10^{-11}	4.25×10^{-6}	5.92×10^{2}	3.59×10^{3}
b	2.81×10^{-10}	1.68×10^{-5}	6.10×10^{2}	3.75×10^{3}
c	1.03×10^{-6}	7.04×10^{-4}	5.97×10^{2}	3.48×10^{3}
d	9.80×10^{-7}	1.73×10^{-3}	5.94×10^{2}	3.64×10^{3}

We consider three aquifer pumping scenarios. The first scenario is of zero pumping rate, resulting in a stationary hydraulic head field. The other two

Table 4. Penalty Results for Test Case 1, Pumping Scenario 3, Computed Using the Inversion Parameters of Table 2

Set	$\|\|\mathbf{Pu}-\mathbf{s}\|\|_2^2/2$	$\|\|\mathbf{Qv}-\mathbf{t}\|\|_2^2/2$	TV $\|\|\mathcal{D}(\mathbf{u})\|\|_1$	$\|\|\mathcal{D}(\mathbf{u})\|\|_2^2$
a	2.24×10^{-10}	2.13×10^{-5}	1.06×10^{3}	1.52×10^{4}
b	6.01×10^{-8}	1.85×10^{-3}	8.72×10^{2}	9.18×10^{3}
c	5.95×10^{-7}	1.68×10^{-3}	6.38×10^{2}	5.78×10^{3}
d	1.07×10^{-6}	5.35×10^{-3}	6.06×10^{2}	4.29×10^{3}

scenarios represent a single pumping location of rate and location indicated in Table 1, i.e., synthetic pumping tests. The pumping locations are shown in Figures 3 and 4. Scenario 2 corresponds to a pumping well located at the low-conductivity region, so that there is significant observed drawdown only in the immediate vicinity of the pumping location. Scenario 3 corresponds to a pumping well located at the high-conductivity region, so that there is significant observed drawdown at various observation locations through the high- and low-conductivity regions. The pumping rate for Scenario 3 is chosen so that observed drawdowns are comparable to that of Scenario 2.

Measurements of hydraulic head are taken at $N_{obs,h}=25$ locations and at $N_t=6$ discrete times $(10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1})$ $[T]$. Given the initial rapid increase in drawdown at the observation locations, we employ a logarithmic scale for the observation times in order to capture this behavior.

A first-order finite element scheme is used to compute the head response at the nodes of each conductivity cell given internal (water discharge or aquifer pumping) and boundary conditions. Boundary conditions are of constant hydraulic heads $h=20$ $[L]$ along the left boundary and $h=10$ $[L]$ along the right boundary. No-flow boundary conditions are imposed along the upper and lower boundaries. The initial condition is the stationary hydraulic head field without aquifer pumping.

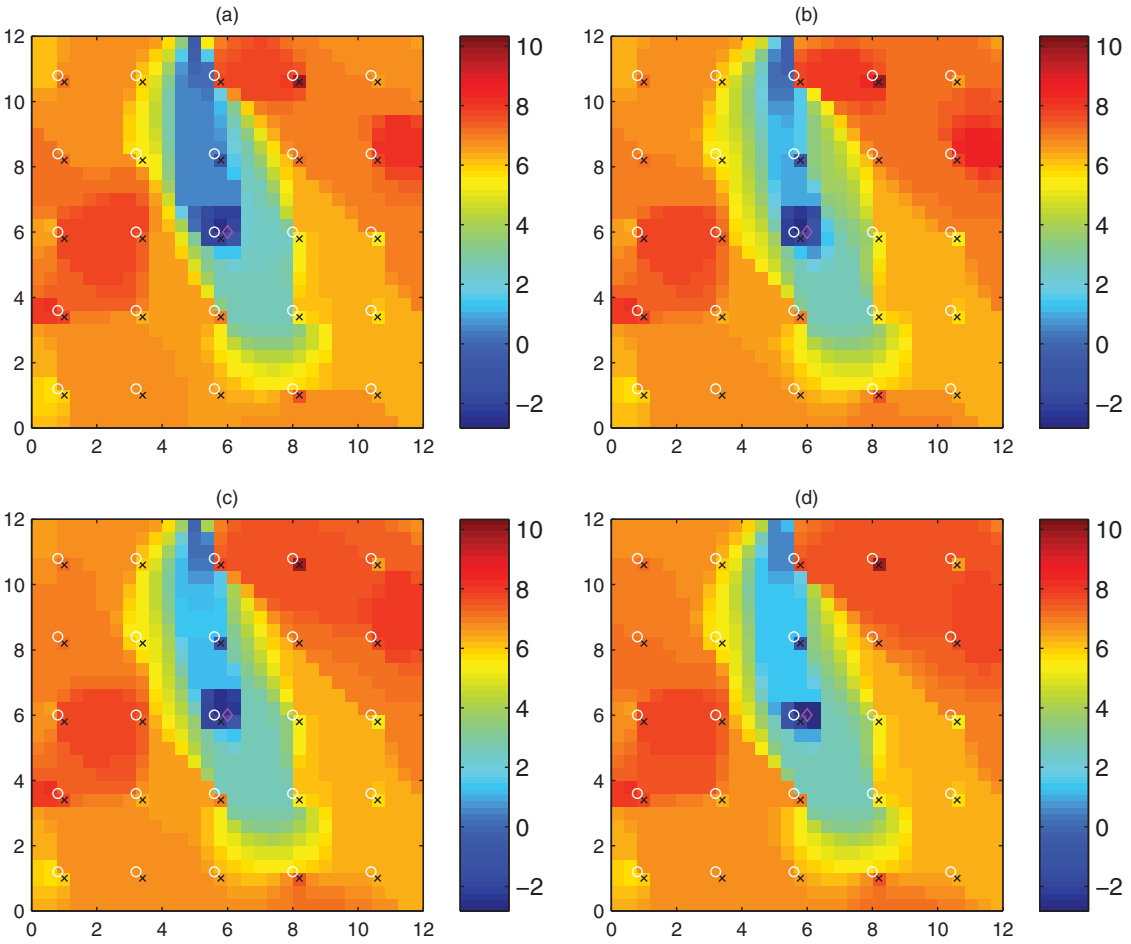

Figure 3. MAP estimator of hydraulic conductivity field for test case 1, Scenario 2, computed using the (a)–(d) inversion parameter set in Table 2.

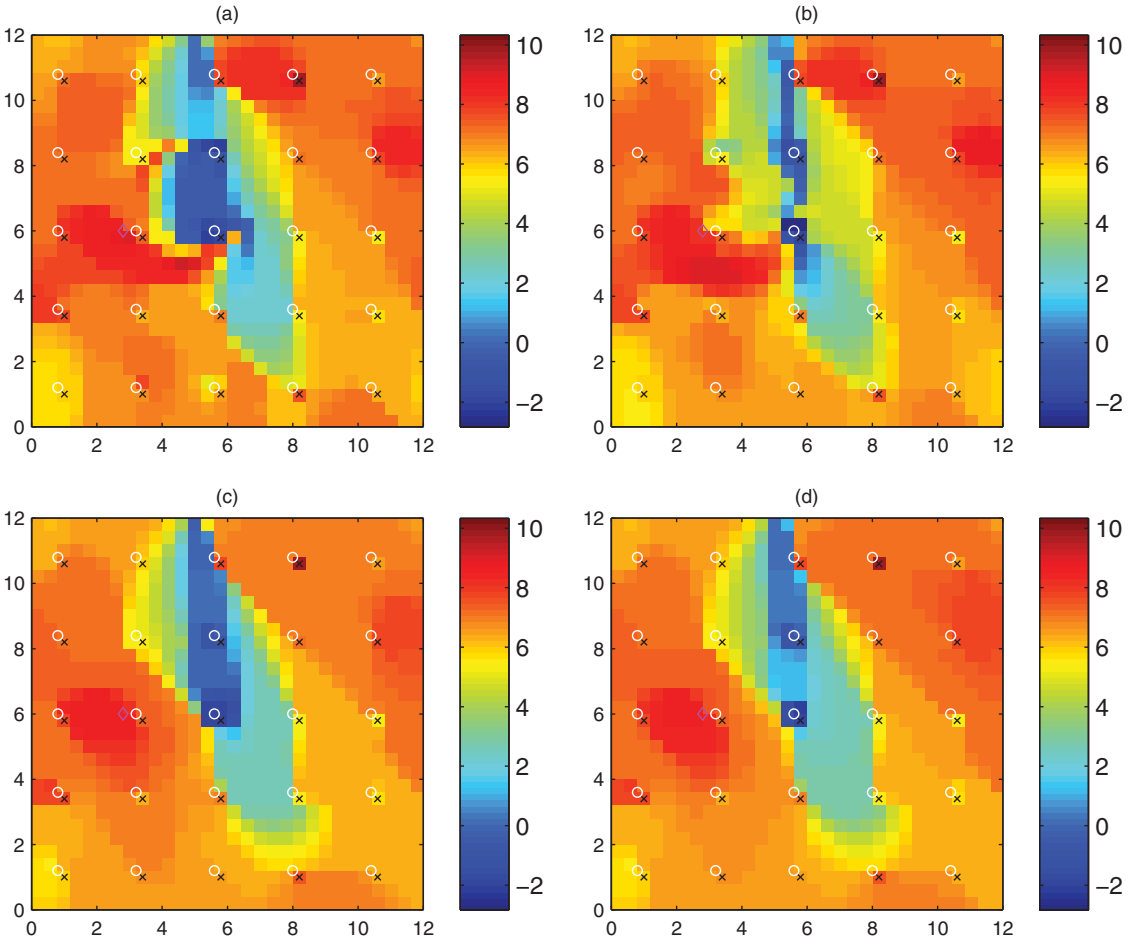

Figure 4. MAP estimator of hydraulic conductivity field for test case 1, Scenario 3, computed using the (a)–(d) inversion parameter set in Table 2.

Additionally, the inversion is performed for four combinations of the inversion parameters α, β, γ, and δ of equation (3), indicated in Table 2. The parameter sets are chosen to evaluate the effect of different degrees of penalization of the deviation of reconstructed heads, TV regularization, and ℓ_2 regularization.

Figure 2 shows the estimated conductivity fields obtained for the first (pumping-free) scenario, together with the set (a) of inversion parameters (as listed in Table 2). The MAP estimator captures the properties of the discrete structures of the original field, namely, the location, orientation and extent of the low-conductivity intrusion, even though the intrusion itself is highly undersampled (2 out of 25 conductivity measurements). Additionally, the intrusion is reconstructed as a region of conductivity higher than in the original field.

The satisfactory reconstruction is a result of the information provided by the hydraulic head observations, together with the piecewise a priori information provided by the TV norm. One can think of the heads data fidelity term, after the linearization of the functional $\mathbf{h}=\mathbf{f}(\mathbf{k})$ around the solution, as a quadratic form

$$\frac{1}{2}\mathbf{w}^\top\mathbf{A}^\top\mathbf{Q}^\top\mathbf{Q}\mathbf{A}\mathbf{w}+(\text{other terms}).$$

The rank of the symmetric positive semidefinite matrix $\mathbf{A}^\top\mathbf{Q}^\top\mathbf{Q}\mathbf{A}$ gives an idea of the dimensionality of the parameter subspace observable via measurements of the system state. In the first pumping scenario the matrix $\mathbf{A}^\top\mathbf{Q}^\top\mathbf{Q}\mathbf{A}$, evaluated around the estimated field shown in Figure 2, has an effective rank of 25 (i.e., equal to the number of head observations). Its eigenvectors show that the introduction of state measurements allows for observation of features that extend spatially beyond the observation points. This added

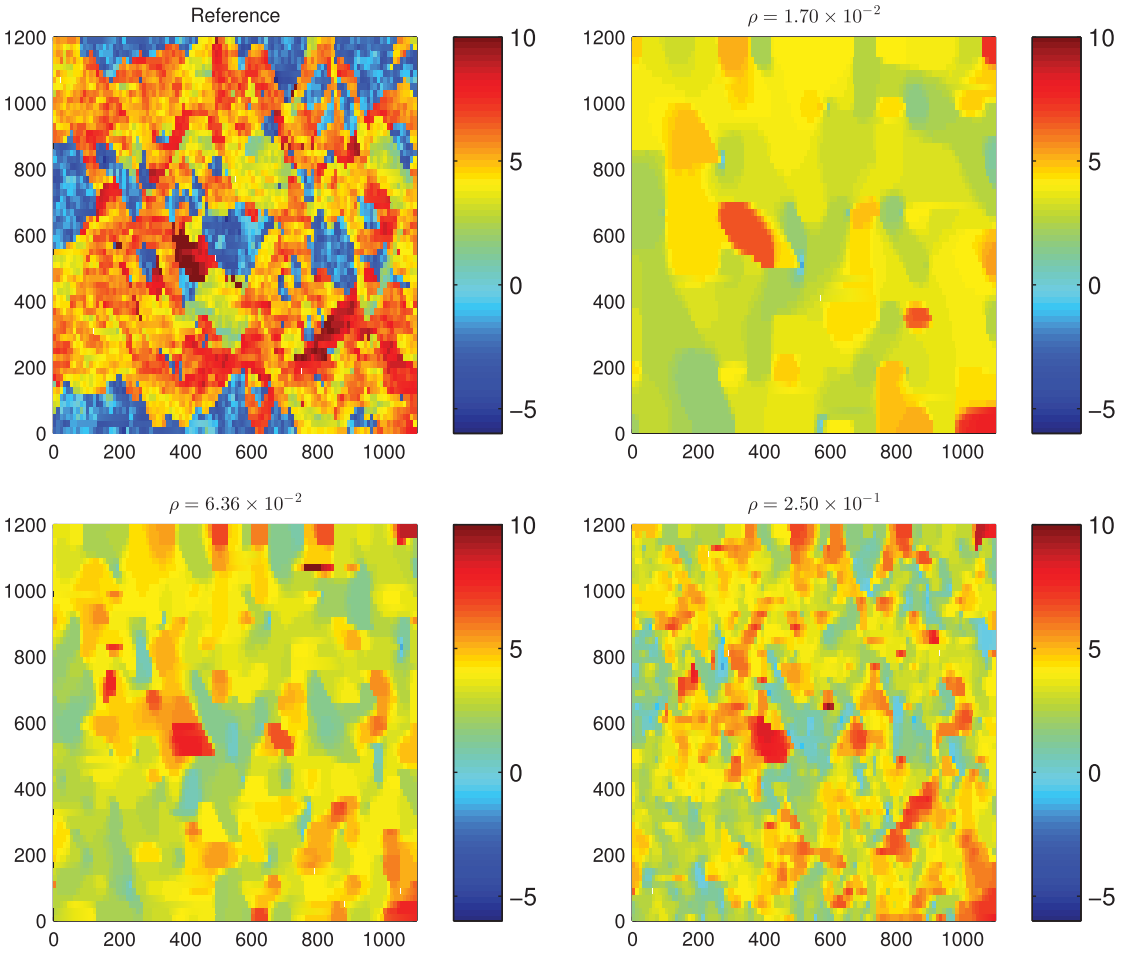

Figure 5. Reference log hydraulic conductivity field and MAP estimators for test case 2, computed for various densities ρ of the observation network of hydraulic head, $\gamma = 1 \times 10^{-8}$, and $\delta = 1 \times 10^{-10}$.

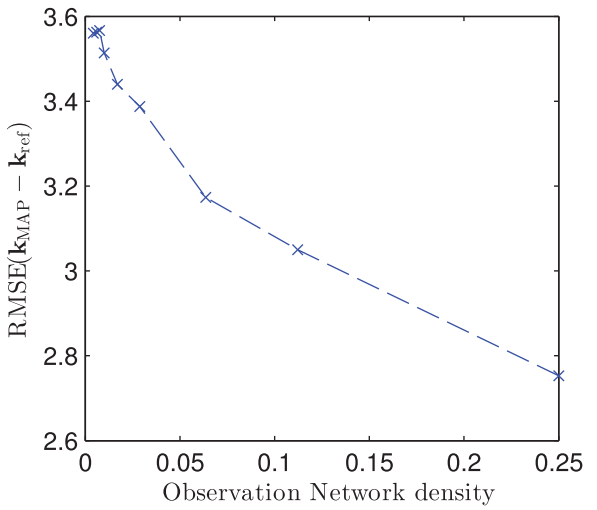

Figure 6. RMSE of MAP estimator of hydraulic conductivity field for test case 2, for varying density of the observation network of hydraulic head.

information, together with the piecewise regularity assumptions encapsulated into the TV norm, allows for the reconstruction shown in Figure 2.

While the reconstruction in Figure 2 captures the overall trend in hydraulic conductivity, it treats the actual observed cells as isolated, pointwise features. This is an effect of the TV norm, which in dimensions $n > 1$ penalizes discontinuities along $(n-1)$-dimensional manifolds separating facies, in favor of point discontinuities. Point discontinuities imply only n jumps per discontinuity and therefore introduce less total variation than a jump across a manifold of large surface area. This effect is more noticeable at the observed cells with values more separated from the overall trend of the field.

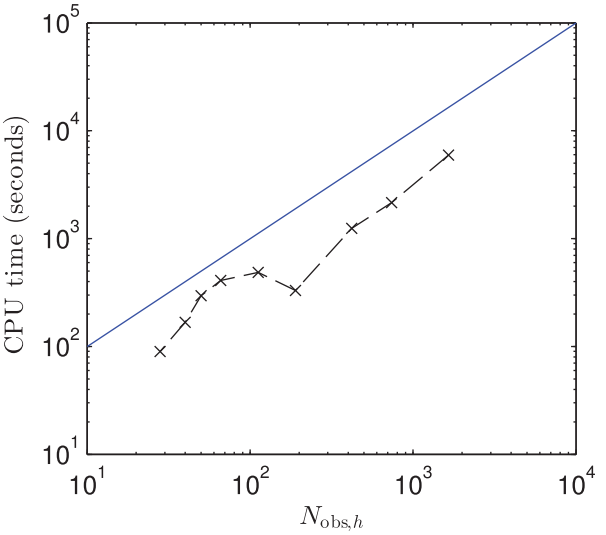

Figure 7. CPU time of MAP estimator for test case 2, for varying number of hydraulic head observation locations, $N_{\mathrm{obs},h}$. Solid blue line is of unit slope, indicating linear behavior.

Figure 3 shows the estimated conductivity fields obtained for the second pumping scenario. The conductivity estimator is similar to the one obtained for the steady case, albeit with an additional feature. In the immediate vicinity of the head and conductivity observation locations along the low-conductivity intrusion, the conductivity field is more accurately estimated as being of lower conductivity. This effect is more noticeable in the vicinity of the head observation location $(5.6, 6.0)$, for which there is a significant observed drawdown. The extent of the low-conductivity patch around observation point $(5.6, 6.0)$ is limited by the TV regularization. As mentioned earlier, it has a tendency to reduce the area of the surface separating the dark blue patch around $(5.6, 6.0)$ and the surrounding low-conductivity intrusion. Indeed, the extent of this patch is reduced as the TV coefficient γ increases (Figures 3c and 3d). The selection of β and γ (Figure 3) affects the overall regularity of the field and the accuracy of the match of head observations, as seen in Table 3.

Comparison of Figures 2 and 3 reveal that the addition of transient information clearly has an effect on the reconstruction results. This is to be expected because more information is available in Scenario 2. The rank of $\mathbf{A}^\top\mathbf{Q}^\top\mathbf{QA}$, evaluated around the estimated field shown in Figure 3a, increases from 25 (Test 1) to 76, which implies that more information from the underlying field can be recovered.

Figure 4 shows the estimated conductivity fields obtained for the third pumping scenario. Locating the pumping well at the high-conductivity region produces a significant variation of heads in observation locations both along the low-conductivity intrusion and the center and upper left portion of the high-conductivity region. As expected, the addition of more transient data increases the amount of recoverable information. The rank of $\mathbf{A}^\top\mathbf{Q}^\top\mathbf{QA}$ (with $\mathbf{A}$ evaluated around the reconstructed field shown in Figure 4a) increases from 25 (Test 1) to 121.

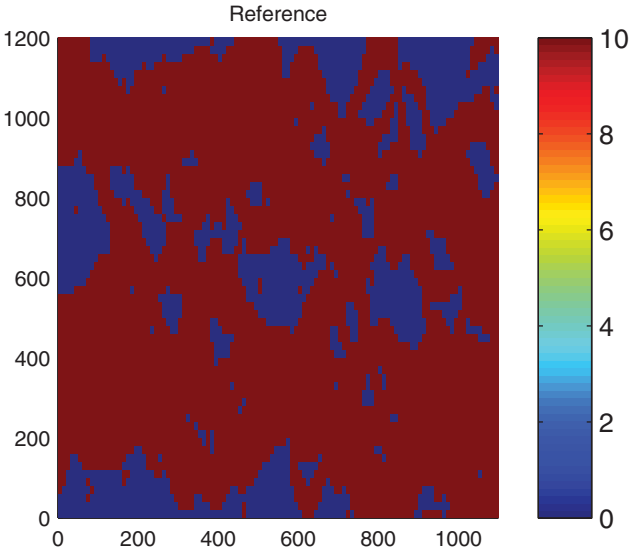

Figure 8. Reference piecewise uniform log hydraulic conductivity field and MAP estimators for test case 3.

Figure 4a shows that for a stringent penalty on the reconstruction of heads ($\beta = 1$) and a relatively lax TV requirement ($\gamma = 1\times10^{-6}$) this increase of information produces a reconstruction of the intrusion that includes two fingers of high conductivity connecting two observation locations. This implies that in order to accurately match head observations (see Table 4 with $\frac{1}{2}||\mathbf{Qv}-\mathbf{t}||_2^2 = 2.13\times10^{-5}$) without introducing either the required heterogeneity into the field (which would significantly increase TV) or the well-delineated border between facies (which is not observable), our reconstruction approach prefers to pay a small price in TV by introducing the aforementioned fingers of high conductivity. It is to be expected that a more regular reconstruction can be obtained either by relaxing

Table 5. Penalty Results for the Piecewise Uniform Synthetic Field of Figure 8

Set	ρ	α	β	γ	δ	$\|\|\mathbf{Qv}-\mathbf{t}\|\|_2^2/2$	RMSE ($\mathbf{k}_{ref}-\mathbf{k}_{MAP}$)
a	2.88×10^{-2}	0.0	1.0	1.0×10^{-8}	1.0×10^{-10}	5.08×10^{-8}	6.470
b	2.88×10^{-2}	0.0	1.0	0.0	2.0×10^{-8}	1.61×10^{-8}	6.478

the penalty on reconstruction of heads (Figure 4b, which increases $\frac{1}{2}||\mathbf{Qv}-\mathbf{t}||_2^2$ from 2.13×10^{-5} to 1.85×10^{-3}) or by increasing the TV coefficient (Figure 4c, which increases $\frac{1}{2}||\mathbf{Qv}-\mathbf{t}||_2^2$ to 1.68×10^{-3}, but decreases TV from 1.06×10^{3} to 6.38×10^{2}). A proper selection of inversion parameters reflecting the desired balance between reconstruction and regularization is the responsibility of the modeler.

6.1. Computational Effort

In addition to the test case studied above, we consider a second test case characterized by a more complex heterogeneity pattern, shown in Figure 5. The reference hydraulic conductivity field is extracted from the bottom layer of the SPE10 test case, discretized on a 60 $\times$ 110 grid of elements of size 20×10 [L]. Similar to Figure 2, we compute the MAP estimator from stationary observations of hydraulic head. No observations of hydraulic conductivity are taken, which corresponds to $\alpha = 0$. Boundary conditions for hydraulic head are again 20 [L] and 10 [L] along the left and right boundaries, respectively, and no-flow along the top and bottom boundaries.

In order to evaluate the scaling properties of the proposed methodology with problem size, we compute the MAP estimator for hydraulic head observation networks of varying density. The density of the observation network ρ is defined as the ratio of observation locations to total hydraulic head degrees of freedom, i.e., $\rho = N_{obs,h}/N_h$. The MAP estimators for various observation network densities are shown in Figure 5.

Figure 6 shows the RMS difference between the MAP estimator $\mathbf{k}_{MAP}$ and the reference field $\mathbf{k}_{ref}$ for increasing density ρ. Together with Figure 5 it demonstrates that, as the density increases, the heterogeneity pattern of the reference field is captured more accurately. As the minimizer of the functional (3), the MAP estimator $\mathbf{k}_{MAP}$ approaches, with increasing density, a hydraulic conductivity configuration that minimizes the total variation, while reconstructing hydraulic head observations accurately.

In order to evaluate the computational effort as a function of problem size, we present in Figure 7 the CPU time of computing $\mathbf{k}_{MAP}$ for different values of $N_{obs,h}$. For this problem, the dominant computational cost is that of the inner iterations (7)–(18), particularly the solution of the linear system (13), which we solve employing the procedure outlined in section 5.2. Figure 7 shows shown that CPU time scales approximately linearly with increasing $N_{obs,h}$, which indicates that the computational cost of the solution to the dense linear system (34) scales linearly with $N_{obs,h}$. This is also consistent with the linear scaling with $N_{obs,h}$ of the

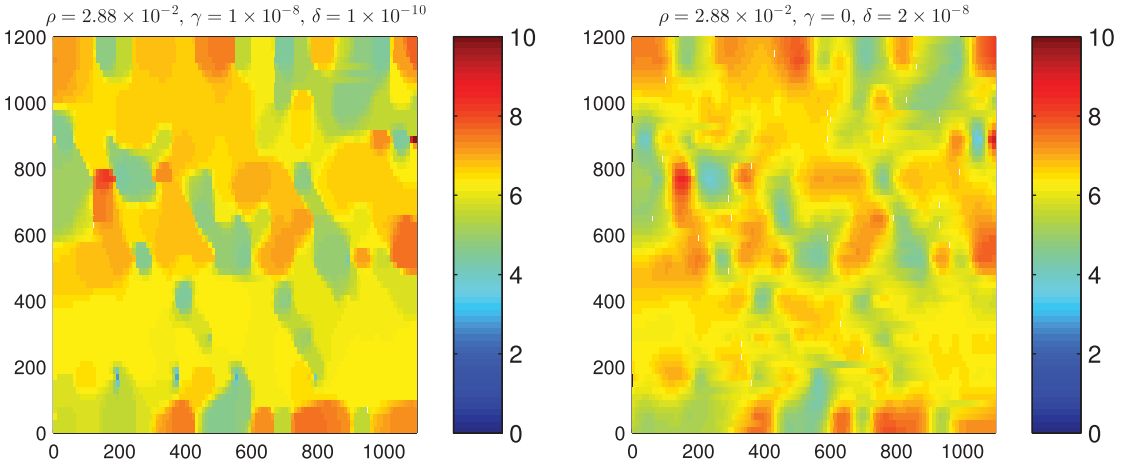

Figure 9. MAP estimators of the reference conductivity field 8, computed using a hydraulic head observation network with density $\rho = 2.88\times10^{-2}$.

computational cost of computing the Jacobian matrix via the adjoint sensitivity analysis (31) discussed in section 5.1.

Finally, in order to emphasize the usefulness of the TV norm for parameter estimation, we consider the synthetic hydraulic conductivity field in Figure 8. This field is obtained from the reference field of Figure 5 by setting all conductivity values larger than 0 to 10, and less or equal to 0 to -6. Using a hydraulic head observation network with density $\rho=2.88\times10^{-2}$, we compute the MAP estimator using the inversion parameters shown in Table 5. The set (b) of inversion parameters corresponds to computing the estimator without TV regularization, and only L^2 regularization. The MAP estimators are shown in Figure 9.

Table 5 shows that the MAP estimators computed with and without TV regularization both accurately recover hydraulic head observations and have an approximately equal RMS difference with respect to the reference field. Nevertheless, Figure 9 shows that incorporating TV regularization allows the modeler to more clearly identify the discrete structures of the hydraulic conductivity field. Not employing TV regularization on the other hand results in a smoother estimator, with less clearly separated structures.

7. Conclusions and Future Work

Our numerical experiments show that the linearized functional minimization algorithm is a feasible and promising approach for inverse modeling of geophysical systems. The applications to a steady flow inversion problem show that the strategy can be used to detect large-scale discrete structures of the parameter field, provided it is equipped with appropriate regularization terms reflecting the prior knowledge of the field (e.g., the TV norm). Incorporation of transient information about the state variables increases the amount of information that can be recovered (Scenarios 2 and 3 shown in Figures 3 and 4). The modeler must pay attention to a possible appearance of extraneous features in the reconstructed field. These are introduced by the algorithm in order to match additional head observations, while keeping the total variation of the field low (Figure 4).

The approach is not limited to systems described by linear governing partial differential equations, such as the groundwater flow equation (1). All that is required is for the Jacobian of the implicit functional relating the system parameters to the system states to be computable. This implies that the strategy can be easily expanded to incorporate measurements of multiple state variables and both measurements and previous knowledge of multiple spatially distributed parameters.

Our linearized functional minimization approach to inverse modeling presents a promising avenue of research. Future work will be focused on (i) the inclusion of other sources of data, such as quantitative transport observations (e.g., tracer data, arrival times, and breakthrough curves), or qualitative or "soft" data; and (ii) the design of a regularization operator that improves on TV by addressing its observed limitations. These limitations include the tendency to penalize large discontinuity surfaces in favor of smaller surfaces or even point discontinuities (see section 6). Additionally, the trust region approach used to minimize (6) can be improved for robustness and efficiency, together with the selection of the inversion parameters.

Acknowledgments

This research was supported in part by the Environmental Programs Directorate of the Los Alamos National Laboratory, the Advanced Simulation Capability for Environmental Management (ASCEM; Department of Energy) project, Air Force Office of Scientific Research (DE-FG02–07ER25815), and National Science Foundation (EAR-1246315). All of the data used in the development of the manuscript are generated from sources expressed in the reference list.

References

Ascher, U., and E. Haber (2004), Computational methods for large distributed parameter estimation problems with possible discontinuities, paper presented at Symposium on Inverse Problems, Design &Optimization, Natl. Sci. Found., Rio de Janeiro, Brazil.

Bachmayr, M., and M. Burger (2009), Iterative total variation schemes for nonlinear inverse problems, *Inverse Probl.*, *25*(10), 105004.

Boyd, S., N. Parikh, E. Chu, B. Peleato, and J. Eckstein (2010), Distributed optimization and statistical learning via the alternating direction method of multipliers, *Found. Trends Mach. Learning*, *3*(1), 1–122, doi:10.1561/2200000016.

Bube, K. P., and R. T. Langan (1997), Hybrid ℓ^1/ℓ^2 minimization with applications to tomography, *Geophysics*, *62*(4), 1183–1195, doi:10.1190/1.1444219.

Cao, Y., S. Li, L. Petzold, and R. Serban (2003), Adjoint sensitivity analysis for differential-algebraic equations: The adjoint DAE system and its numerical solution, *SIAM J. Sci. Comput.*, *24*(3), 1076–1089.

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

Chung, E. T., T. F. Chan, and X.-C. Tai (2005), Electrical impedance tomography using level set representation and total variational regularization, *J. Comput. Phys.*, *205*(1), 357–372, doi:10.1016/j.jcp.2004.11.022.

de Marsily, G., F. Delay, J. Gonalvs, P. Renard, V. Teles, and S. Violette (2005), Dealing with spatial heterogeneity, *Hydrogeol. J.*, *13*(1), 161–183, doi:10.1007/s10040-004-0432-3.

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

Goldstein, T., and S. J. Osher (2009), The Split Bregman method for L1-regularized problems, *SIAM J. Imaging Sci.*, *2*(2), 323–343, doi:10.1137/080725891.

Hairer, E., and G. Wanner (1999), Stiff differential equations solved by Radau methods, *J. Comput. Appl. Math.*, *111*(1–2), 93–111, doi:10.1016/S0377-0427(99)00134-X.

Hu, L. (2002), Combination of dependent realizations within the gradual deformation method, *Math. Geol.*, *34*(8), 953–963, doi:10.1023/A:1021316707087.

Hu, L., and M. Le Ravalec-Dupin (2004), An improved gradual deformation method for reconciling random and gradient searches in stochastic optimizations, *Math. Geol.*, *36*(6), 703–719, doi:10.1023/B:MATG.0000039542.73994.a2.

Iglesias, M. A., and D. McLaughlin (2011), Level-set techniques for facies identification in reservoir modeling, *Inverse Probl.*, *27*, 035008, doi:10.1088/0266-5611/27/3/035008.

Lee, J., and P. K. Kitanidis (2013), Bayesian inversion with total variation prior for discrete geologic structure identification, *Water Resour. Res.*, *49*, 7658–7669, doi:10.1002/2012WR013431.

Nœtinger, B., V. Artus, and G. Zargar (2005), The future of stochastic and upscaling methods in hydrogeology, *Hydrogeol. J.*, *13*(1), 184–201, doi:10.1007/s10040-004-0427-0.

Rodríguez, P., and B. Wohlberg (2009), Efficient minimization method for a generalized total variation functional, *IEEE Trans. Image Processing*, *18*(2), 322–332, doi:10.1109/TIP.2008.2008420.

Rudin, L., S. J. Osher, and E. Fatemi (1992), Nonlinear total variation based noise removal algorithms, *Physica D*, *60*(1–4), 259–268, doi:10.1016/0167-2789(92)90242-F.

Serban, R., and A. C. Hindmarsh (2005), Cvodes, the sensitivity-enabled ode solver in sundials, paper presented at 5th International Conference on Multibody Systems, Nonlinear Dynamics and Control, Design Eng. Div. and Comput. and Inf. in Eng. Div., Long Beach, Calif.

Tartakovsky, D. M., and B. Wohlberg (2004), Delineation of geologic facies with statistical learning theory, *Geophys. Res. Lett.*, *31*, L18502, doi:10.1029/2004GL020864.

Winter, C. L., D. M. Tartakovsky, and A. Guadagnini (2003), Moment differential equations for flow in highly heterogeneous porous media, *Surv. Geophys.*, *24*(1), 81–106.

Wohlberg, B., D. M. Tartakovsky, and A. Guadagnini (2006), Subsurface characterization with support vector machines, *IEEE Trans. Geosci. Remote Sens.*, *44*(1), 47–57, doi:10.1109/TGRS.2005.859953.

Wohlberg, B., D. M. Tartakovsky, and M. Dentz (2012), Linearized functional minimization for inverse modeling, paper presented at *XIX International Conference on Water Resources (CMWR)*, Univ. of Ill. at Urbana-Champaign, Urbana-Champaign.