

A Markov random field model of contamination source identification in porous media flow

Jingbo Wang, Nicholas Zabaras *

Materials Process Design and Control Laboratory, Sibley School of Mechanical and Aerospace Engineering, 188 Frank H.T. Rhodes Hall, Cornell University, Ithaca, NY 14853-3801, USA

Received 7 December 2004; received in revised form 20 June 2005
Available online 21 November 2005

Abstract

A contamination source identification problem in constant porous media flow is addressed by solving the advection–dispersion equation (ADE) with a hierarchical Bayesian computation method backward through time. The contaminant concentration is modeled as a pair-wise Markov random field (MRF) and the distribution is updated using current concentration measurements at finite locations. Hierarchical Bayesian analysis is used to derive the posterior distribution of the contaminant concentration at past time points. The posterior mean estimate is computed using a modified single-component Gibbs algorithm. The methodology is first tested via examples of contaminant identification in a homogeneous porous medium using both diffusion-dominated and convection-dominated conditions. A heterogeneous porous media flow case is also examined. In all the numerical studies reported, the anisotropic dispersion effect is considered. It is verified that the MRF model can effectively model the spatial correlation of the concentration field, and the presented approach can provide accurate solutions to the ill-posed inverse problem.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Porous media; Advection; Dispersion; Pollution source; Source identification; Bayesian inference; Inverse problems; Markov random field

1. Introduction

The contamination source identification problem has received significant research interest due to its applications in groundwater and soil cleanup. Addressing this problem requires solving the partial differential equations (PDEs) of contaminant propagation in porous media flow backwards in time. Namely, the objective is to compute the history of contaminant concentration from current concentration data. The ill-posedness of this inverse problem and the difficulties in simulating the contaminant propagation have been well-recognized. To facilitate the solution to this challenging problem, a variety of methods have been developed over the past several decades, which have been reviewed by Atmadja and Bagtzoglou [1] and Michalak and Kitanidis [2].

Among the solution methods considered, Bayesian computation has a number of distinctive attributes. The Bayesian approach restates the ill-posed inverse problem as a well-posed problem in an expanded stochastic space. Instead of computing only deterministic point estimates, it computes the probability distribution (often called posterior distribution) of contaminant concentration history conditional on current concentration data. The Bayesian approach can quantify random errors in the data and uncertainties in the concentration field. The posterior distribution is composed of the likelihood, which is the probability of the data given the concentration history, and the prior distribution, which models the regularity of the unknown concentration field. Incorporation of the prior distribution allows flexible regularization and accurate estimation from sparse data.

The Bayesian approach was introduced for solving contamination source identification by Snodgrass and Kitanidis [3]. In [4,5], Michalak and Kitanidis developed

* Corresponding author. Tel.: +1 607 255 9104; fax: +1 607 255 1222.
E-mail address: zabaras@cornell.edu (N. Zabaras).
URL: <http://mpdc.mae.cornell.edu/> (N. Zabaras).

Nomenclature

A	acceptance probability
c	concentration
\hat{c}	known concentration at well
c_0	initial concentration
\mathbf{D}	dispersion tensor
f	linear FE basis function
H	sensitivity matrix
\mathbf{I}	identity matrix
K	permeability
L	length of Markov chain
L_2	square-integrable function space
m	dimension of θ
N	total number of data
\mathbf{n}	unit normal direction vector
p	pressure, probability density function
P_e	local Peclet number
q	volume flux, rate proposal distribution
t	time
Δt	time step size
T	maximum time of simulation
u	random number
\mathbf{u}	Darcy velocity
U	uniform distribution
w	test function of concentration
\mathbf{w}	test function of velocity
W	precision matrix
\mathbf{x}	spatial vector
Y	concentration data vector

Greek symbols

α	Gamma distribution parameter
α_m	molecular diffusivity

α_l	longitude dispersion coefficient
α_t	transverse dispersion coefficient
β	Gamma distribution parameter
γ	scaling constant of MRF
Γ	Gamma function
θ	parameter form of unknown c
$\hat{\theta}$	estimate of θ
λ	scaling constant of MRF
μ	dynamic viscosity
ξ	parameter of hyper-prior
σ	standard deviation of ω_m
τ	upwind parameter
ϕ	porosity
Φ	kernel function of MRF
ω_m	random measurement error
Ω	spatial domain
Ω_e	finite element domain
$\partial\Omega$	spatial domain boundary

Superscripts

T	transpose
(i)	i th iteration or i th time step
(*)	candidate

Subscripts

i	i th component
$-i$	full conditional
$i \sim j$	site neighborhood

a confined Brownian motion model to enforce nonnegativity of concentration estimates and to select structure parameters of the Bayesian posterior distribution. In our work, a pair-wise Markov random field is introduced to regularize the prior distribution of concentration history. A hierarchical Bayesian analysis is introduced, allowing the uncertainty (structure) parameters to be quantified as part of the posterior distribution and to be explored using a mixed Metropolis–Hastings (MH) algorithm [6].

In contrast to the relatively straightforward formulation of the likelihood, modeling the prior distribution in the Bayesian approach is more difficult. Considering the structure of a spatially varying concentration field, a pair-wise MRF model is introduced from spatial statistics [7]. The MRF models the spatial correlation among concentration variables at adjacent locations, which smoothes the concentration field. This model has been used in a variety of spatial applications [8,9]. Very

recently, the model was introduced to inverse heat transfer problems by Wang and Zabarar [10,11], where time is treated as an extra spatial dimension in modeling dynamic fields. In [11], the canonical MRF model was enhanced to model random fields with different temporal and spatial length scales and to resolve discontinuities in the unknown fields.

In this work, in addition to the concentration field, the standard deviation of measurement errors and the scaling parameter of the prior distribution are also treated as random variables. We refer to these parameters as ‘structure variables’ following the terminology by Michalak and Kitanidis [4]. These parameters are often called hyperparameters in Bayesian statistics. Hierarchical Bayesian analysis [12] is used to derive the joint distribution of structure parameters with the unknown concentration field. The joint posterior state space is then explored using a mixed Markov chain Monte Carlo (MCMC) sampler that samples the concentration variables using the Gibbs algorithm

[13] and the structure parameters using the MH algorithm [14].

Simulation data are used in this study to test the presented inverse computation method. The Darcy equation for porous media flow is first solved using the global gradient post-processing method [15]. The velocity field is then used to solve the advection–dispersion equation of concentration using a streamline-upwind/Petrov–Galerkin (SUPG) finite element method. All equations are solved on a rather fine grid to generate simulation data, thus avoiding the so called ‘inverse crimes’ [16].

The rest of this article is organized as follows. Section 2 introduces the mathematical definition of the problem. The direct simulation of the contamination propagation is discussed in Section 3. It is then followed by the hierarchical Bayesian formulation of the inverse computation. The posterior exploration algorithms are presented in Section 5. Section 6 contains numerical examples to demonstrate the developed methodology. Finally, conclusions of the current work and future research issues are summarized in Section 7.

2. Problem definition

Propagation of contaminant in an impermeable porous medium can be described by the following advection–dispersion equation (ADE) [17]:

$$\phi \frac{\partial c}{\partial t} + \nabla \cdot (c\mathbf{u}) - \nabla \cdot (\mathbf{D}\nabla c) = \hat{c}q, \quad \text{in } \Omega \times (0, T], \quad (1)$$

with prescribed initial and (Neumann) boundary conditions,

$$c(\mathbf{x}, 0) = c_0(\mathbf{x}), \quad \text{in } \Omega, \quad (2)$$

$$\mathbf{D}\nabla c \cdot \mathbf{n} = 0, \quad \text{on } \partial\Omega \times (0, T]. \quad (3)$$

In the above equations, c is the concentration (mass fraction) of the contaminant and \hat{c} is the prescribed concentration values at the injection and production wells. Also q denotes the volume flux rate at the wells and ϕ and \mathbf{D} are the medium porosity and dispersion tensor, respectively. Finally, Ω is the spatial domain and $(0, T]$ is the total time span. The anisotropic dispersion coefficient \mathbf{D} can be modeled as follows:

$$\mathbf{D} = \phi \{ \alpha_m \mathbf{I} + \|\mathbf{u}\| [\alpha_l E(\mathbf{u}) + \alpha_t E^\perp(\mathbf{u})] \}, \quad (4)$$

with

$$E(\mathbf{u}) = \frac{1}{\|\mathbf{u}\|^2} \mathbf{u} \otimes \mathbf{u}, \quad E^\perp(\mathbf{u}) = \mathbf{I} - E(\mathbf{u}), \quad (5)$$

where \mathbf{I} is the identity matrix, and α_m , α_l and α_t are the molecular diffusivity, longitudinal dispersion coefficient and transverse dispersion coefficient, respectively. The Darcy velocity \mathbf{u} can be computed from the following equations:

$$\nabla \cdot \mathbf{u} = q, \quad \text{in } \Omega \times (0, T], \quad (6)$$

$$\mathbf{u} = -\frac{K(\mathbf{x})}{\mu(c)} \nabla p, \quad \text{in } \Omega \times (0, T], \quad (7)$$

$$\mathbf{u} \cdot \mathbf{n} = 0, \quad \text{on } \partial\Omega \times (0, T], \quad (8)$$

where p is the hydrodynamic pressure and K and μ are the permeability and dynamic viscosity, respectively. In this study, we assume the variation of viscosity can be neglected, i.e. μ is a constant equal to the dynamic viscosity of the resident fluid (water). Therefore, the ADE equation (1) is decoupled from the flow Eqs. (6) and (7).

A direct (or forward) contaminant propagation problem is defined as the computation of the concentration distribution at all times $t \in (0, T]$, given initial condition equation (2) and boundary condition equation (3). In the inverse problem of interest considered, the contamination concentration at current time $t = T$ can be measured at finite locations inside Ω . However, the history of contaminant distribution is not known. Namely, c_0 and the time span T between releasing time $t = 0$ of the contaminant and measurement time $t = T$ are both unknown. The inverse problem is to compute the concentration backward in time, namely $c(t)$ with $t < T$, on a finer scale grid than the measurement scale grid. It is assumed that no prior knowledge of releasing time and location of the contaminant is available. The releasing time is defined as the time point corresponding to a backward computed concentration of 1.0 at any location.

3. The direct simulation and sensitivity analysis

3.1. Solution to the flow equations

Solution to the direct problem is required for the inverse computation. The direct simulation can be separated into two parts: solution to the flow equations and solution to the concentration equation. In the first part, the constant flow velocity field is obtained by solving Eqs. (6)–(8). The velocity is then used in solving Eqs. (1)–(5).

In the context of the finite element (FE) method, the most common approaches to solving the flow equations (6)–(8) are the stabilized finite element method, in which the pressure and velocity are determined simultaneously, and the gradient post-processing method, in which the pressure is found first and then the velocity is calculated via gradient post-processing. The gradient post-processing method is easier to implement and computationally less costly. It solves a diffusion equation derived by substituting Eq. (7) into Eq. (6) for pressure first. The velocity is then computed as the smoothed gradient of the pressure field. In this work, the flow equations are solved using a global post-processing method as discussed in [15].

In the gradient post-processing approach, the pressure is solved using

$$\nabla \cdot \left(\frac{K}{\mu} \nabla p \right) = -q, \quad (9)$$

which is derived by substituting Eq. (7) into Eq. (6). The finite element technique to solve this steady state diffusion equation is trivial. Once the pressure field is obtained, Eq. (7) can be used to compute the velocity. However, since the finite element solution of pressure is usually not smooth across element boundaries, the velocity obtained by directly computing the gradient of pressure is discontinuous across element boundaries, which is not physically feasible. To achieve a continuous velocity solution, a global L_2 -smoothing post-processing problem is usually solved with the following weak formulation:

$$(\mathbf{u}, \mathbf{w}) = \left(-\frac{K}{\mu} \nabla p, \mathbf{w} \right), \tag{10}$$

where (\cdot, \cdot) is the $L_2(\Omega)$ inner product and \mathbf{w} is the test function for velocity. According to Loula et al. [15], to further increase the accuracy of the post-processing result, Eq. (10) is often modified as:

$$(\mathbf{u}, \mathbf{w}) + (\delta h)^\alpha (\nabla \cdot \mathbf{u}, \nabla \cdot \mathbf{w}) = \left(-\frac{K}{\mu} \nabla p, \mathbf{w} \right) + (\delta h)^\alpha (q, \nabla \cdot \mathbf{w}), \tag{11}$$

in which h is the finite element grid size. The parameters δ and α are here taken as 0.1 and 1, respectively [15]. Let

$$\mathcal{U} = \{ \mathbf{u} | \mathbf{u} \in (L_2(\Omega))^{\text{dim}}, \nabla \cdot \mathbf{u} \in L_2(\Omega), \mathbf{u} \cdot \mathbf{n} = 0 \}. \tag{12}$$

The problem can be stated as to find $\mathbf{u} \in \mathcal{U}$, such that, for all $\mathbf{w} \in \mathcal{U}$, Eq. (11) holds.

3.2. Solution of the concentration equation

After computing the velocity from the above approaches, one can evaluate the concentration using Eq. (1). To solve this advection–dispersion equation, the SUPG finite element formulation is used [19]:

$$\begin{aligned} & \int_{\Omega} \phi \frac{\partial c}{\partial t} w d\Omega + \int_{\Omega} (\mathbf{u} \cdot \nabla c) w d\Omega + \int_{\Omega} q c w d\Omega \\ & + \int_{\Omega} \mathbf{D} \nabla c \cdot \nabla w d\Omega + \sum_{e=1}^{nel} \int_{\Omega_e} \tau \mathbf{u}_e \cdot \nabla w \left(\phi \frac{\partial c}{\partial t} + \mathbf{u}_e \cdot \nabla c + q c \right) d\Omega_e \\ & = \int_{\Omega} \hat{c} q w d\Omega + \sum_{e=1}^{nel} \int_{\Omega_e} \tau \mathbf{u}_e \cdot \nabla w \hat{c} q d\Omega_e, \end{aligned} \tag{13}$$

where w is the test function for concentration. The weak problem is to find $c \in H_1(\Omega)$ such that, for all $w \in H_1(\Omega)$ Eq. (13) holds. The element based integrals (the 5th term on the left hand side and the 2nd term on the right hand side) are the SUPG stabilizing terms, in which τ is the upwind parameter. In this formulation, it is assumed that the gradient of the test function w is discontinuous across element boundaries. The stabilization parameter τ is computed via the following formula:

$$\tau = \frac{1}{2} \frac{h}{\|\mathbf{u}_e\|} \min \left(\frac{Pe}{3}, 1.0 \right), \tag{14}$$

where Pe is the local (element) Peclet number that is defined as

$$Pe = \frac{1}{2} h \frac{\|\mathbf{u}_e\|^3}{\mathbf{u}_e^T \mathbf{D} \mathbf{u}_e}. \tag{15}$$

With the finite element formulations introduced above, the direct problem is solved using two-dimensional bi-linear finite elements. The simulator was implemented for parallel machines using PETSc [18] and has been tested by comparing the results to solutions of various numerical examples documented in [19–21].

3.3. Sensitivity analysis

A discussion of sensitivity analysis is necessary to improve understanding of the Bayesian formulation. To present the sensitivity analysis, a simpler inverse problem is temporarily considered in this section. By assuming a known releasing time of contaminant, the inverse problem introduced in Section 2 is reduced to the estimation of a spatially varying function $c_0(\mathbf{x})$. This function estimation problem is further transformed into a parameter estimation problem by introducing the following approximation:

$$\hat{c}_0(\mathbf{x}) = \sum_{j=1}^m \theta_j f_j(\mathbf{x}), \tag{16}$$

where $f_j(\mathbf{x})$'s are the linear finite element basis functions and θ_j 's are the nodal values of finite element approximation of c_0 . The problem now is to estimate an unknown m -dimensional vector θ with $\theta(j) = \theta_j$ being the nodal value associated with the j th basis function.

Let $c(\mathbf{x}, T)$ be the concentration at measurement time $t = T$ that is computed from Eq. (1) using \hat{c}_0 as initial condition. Due to the linearity of the direct problem,

$$c(\mathbf{x}, T) = \sum_{j=1}^m \theta_j c^j(\mathbf{x}, T), \tag{17}$$

with $c^j(\mathbf{x}, T)$ being the direct solution of concentration at $t = T$ using $f_j(\mathbf{x})$ as the initial condition.

Let the N -dimensional vector Y denote the concentration measurement data at $t = T$ with $Y(i)$ being the measurement at the i th sensor location (\mathbf{x}_i) and N being the total number of sensors. Furthermore, let C be the N -dimensional vector with $C(i) = c(\mathbf{x}_i, T)$. Using Eq. (17), C can be represented as

$$C = H\theta, \tag{18}$$

in which H is a $N \times m$ matrix with $H(i, j) = c^j(\mathbf{x}_i, T)$. H is often called the sensitivity matrix. Its component $H(i, j)$ reflects the sensitivity of the concentration $C(i)$ at each sensor location i with respect to small variations in the parameter $\theta(j)$.

In the remaining part of this article, the following system relationship is assumed:

$$Y = H\theta + \omega, \tag{19}$$

where ω denotes the error between the measured data and the concentration computed using the true initial condition. Therefore, ω contains both the random measurement

error and the numerical error. The objective of the simpler inverse problem is to find an estimate of θ such that the discrepancy between Y and C is minimized in some sense.

With the capability to simulate the direct problem and compute the sensitivity matrix, we are now ready to investigate the Bayesian formulation.

4. Bayesian backward computation

4.1. Bayesian inverse formulation

For the parameter estimation problem in Eq. (19), Bayesian inference computes the probability density function of random unknown θ conditional on the data Y using the Bayes' formula:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)}, \tag{20}$$

where the conditional density function $p(Y|\theta)$ is the likelihood and $p(\theta)$ is the prior density function. The conditional density function $p(\theta|Y)$ is often called the posterior probability density. The difference between the Bayesian approach and other inverse methods is that the Bayesian approach determines the distribution of unknown parameters instead of point estimates, and as a result, the inverse problem is formulated as a well-posed problem in a stochastic space (state space defined by the prior distribution). Even in seeking point estimates, the Bayesian approach enables more flexible regularization than other methods. Based upon the posterior distribution, credible intervals at arbitrary levels can be computed for all Bayesian point estimates, which is another advantage of the Bayesian approach.

Despite the fact that a number of point estimates can be computed according to different loss function definitions, the posterior mean estimate (minimum mean square error (MMSE) estimate) is deemed a reliable estimate and is computed in the current study as the inverse solution. The posterior mean estimate is defined as

$$\hat{\theta}_{\text{postmean}} = E\theta|Y. \tag{21}$$

There are two points worth emphasizing here: first, it is only meaningful to study the probability of a random variable existing within an interval, rather than having a particular value; therefore, estimation of the distribution makes more practical sense than computing point estimates. Second, the actual value of the parameter θ is fixed instead of being random. The rationality in modelling it as a random variable is that it is inferred from the noise-polluted data; hence, uncertainties exist in our knowledge of this quantity.

As seen from Eq. (20), to obtain the posterior distribution, one needs to compute $p(Y)$, $p(Y|\theta)$ and $p(\theta)$. However, since Y is the known data, the role of $p(Y)$ is nothing more than a normalizing constant in the posterior distribution. It is not needed to compute the posterior estimates using sampling algorithms introduced later in this article. Therefore, it is enough to evaluate Eq. (20) as follows:

$$p(\theta|Y) \propto p(Y|\theta)p(\theta). \tag{22}$$

If the random errors in Eq. (19) are assumed to be independent identically distributed (i.i.d.) Gauss random noise with zero mean and variance v (standard deviation $\sigma = \sqrt{v}$), the likelihood can be formulated as

$$p(Y|\theta) = \frac{1}{(2\pi)^{N/2}v^{N/2}} \exp \left\{ -\frac{(Y - H\theta)^T(Y - H\theta)}{2v} \right\}. \tag{23}$$

It should be noticed that even though other distributions can be used to model random errors, the Gaussian distribution is the most commonly used model. It is well recognized that in large data set with random errors, the Gaussian distribution fits quite well the actual distribution.

To obtain the posterior distribution, the next task is to formulate the prior distribution $p(\theta)$.

4.2. Markov random field as prior distribution

The prior distribution represents the distribution information of θ in advance of Bayesian inference. It defines a prior state space with a certain degree of regularity that cures the ill-posedness of the inverse problem. There are a number of standard techniques to model the prior distribution, some of which are known as conjugate priors and Jeffery's priors [12].

For the main unknown θ in the current problem, a pair-wise Markov random field model is selected as the prior. The consideration is that θ represents the concentration values at a large set of spatial locations (finite element nodes), while the pair-wise MRF model is designed for modeling the correlation between spatially adjacent random variables on finite lattice system [7]. By modeling the correlation among spatially indexed random variables θ_i 's, regularity is posed on the state space of θ .

The general form of the pair-wise MRF is as follows:

$$p(\theta) \propto \exp \left\{ -\sum_{i \sim j} W_{ij} \Phi(\gamma(\theta_i - \theta_j)) \right\}, \tag{24}$$

where γ is a scaling parameter, $\Phi(\cdot)$ is an even function that determines the specific form of the MRF, the summation is over all pairs of sites $i \sim j$ that are defined as neighbors, and W_{ij} are specified non-zero weights. Let $\Phi(u) = \frac{1}{2}u^2$, the MRF used herein is of the form

$$p(\theta) \propto \lambda^{m/2} \exp \left(-\frac{1}{2} \lambda \theta^T W \theta \right). \tag{25}$$

In the above one-parameter model, the entries of the $m \times m$ matrix W are determined as $W_{ij} = n_i$ if $i = j$, $W_{ij} = -1$ if sites i and j are adjacent (termed as neighbor sites) and as 0 otherwise. The variable n_i denotes the number of neighbors adjacent to site i and λ is a scaling parameter.

There are several benefits of the MRF model equation (25). First of all, it is shift invariant, namely, $p(\theta) = p(\theta + \mathbf{b})$ with \mathbf{b} being a vector with identical components. This ensures that the prior distribution will not over

constrain the unknown parameter. In addition, Eq. (25) is much less computationally expensive to explore than a full-entry Gauss random field. Furthermore, the method is rather flexible in modeling various correlation structures by changing the order of Markov field and weight specification.

It is important to mention that the above pair-wise MRF model can be easily extended to the discontinuity adaptive MRF model that enables automatic resolution of discontinuities in unknown random fields. Interested readers are encouraged to consult our previous study in [11].

With the likelihood Eq. (23) and prior distribution Eq. (25), the posterior can be tentatively written as:

$$p(\theta|Y) \propto \exp\left\{-\frac{(Y - H\theta)^T(Y - H\theta)}{2v}\right\} \cdot \exp\left(-\frac{1}{2}\lambda\theta^T W\theta\right). \tag{26}$$

However, this posterior distribution depends on pre-fixed values of v and λ . In reality, the magnitude of the actual noise can only be roughly estimated. Selection of λ is even more non-deterministic. These two structure parameters are key to the estimation of the posterior distribution and to the degree of smoothness of all point estimates. Unlike earlier methods that try to select such structure parameters before the inverse computation, a hierarchical Bayesian approach is considered here to estimate the distribution of the structure parameters simultaneously with the computation of the concentration distribution.

4.3. The hierarchical posterior distribution

The hierarchical Bayesian posterior is usually used when the prior distribution of the primary unknown variables (e.g. of the concentration history in the current problem) depends on some uncertain parameters, also known as hyper-parameters or structure parameters. The canonical form of a hierarchical posterior can be written as:

$$p(\theta, \xi|Y) \propto p(Y|\theta)p(\theta|\xi)p(\xi), \tag{27}$$

where ξ is the structure parameter and $p(\xi)$ is the prior distribution of ξ . It is assumed in this formulation that the likelihood solely depends on the primary parameter θ . The advantages of using Bayesian hierarchical posterior are: (i) the uncertainty of hyper-parameters is quantified, and (ii) the effect of poor knowledge of structure parameters on the posterior is reduced so that the posterior estimates are less likely to be over-constrained.

The structure parameter is generally assumed to have a nearly non-informative distribution over its support. For instance, in the current example, the structure parameters λ and v are both assumed a priori to be nearly uniformly distributed over $(0, \infty]$. However, the functional form of the nearly non-informative prior varies for different structure parameters. In this study, conjugate priors [12] are used to model the prior distribution of λ and v .

Conjugate priors are the prior distributions that ensure the corresponding posterior distributions having the same

functional form as the priors. For Eq. (26), Gamma and inverse Gamma distributions are chosen as priors for λ and v , respectively:

$$p(\lambda) \propto \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \lambda^{\alpha_1-1} e^{-\beta_1 \lambda}, \quad \lambda \in (0, \infty], \tag{28}$$

$$p(v^{-1}) \propto \frac{\beta_2^{\alpha_2}}{\Gamma(\alpha_2)} v^{-(\alpha_2+1)} e^{-\beta_2 v^{-1}}, \quad v \in (0, \infty], \tag{29}$$

where $\Gamma(\cdot)$ is the standard Gamma function. A small value $1.0e-3$ is selected for the Gamma distribution constants $\alpha_1, \alpha_2, \beta_1$ and β_2 . Thus, the distributions Eqs. (28) and (29) are nearly non-informative over $(0, \infty]$.

With the hyper priors defined above, a hierarchical Bayesian posterior distribution can be computed as follows:

$$\begin{aligned} p(\theta, \lambda, v|Y) &\propto p(Y|\theta, v)p(\theta|\lambda)p(\lambda)p(v) \\ &\propto v^{-N/2} \exp\left\{-\frac{(Y - H\theta)^T(Y - H\theta)}{2v}\right\} \\ &\quad \cdot \lambda^{m/2} \exp\left\{-\frac{1}{2}\lambda\theta^T W\theta\right\} \\ &\quad \cdot \lambda^{\alpha_1-1} \exp\{-\beta_1 \lambda\} v^{-(1+\alpha_2)} \exp\{-\beta_2 v^{-1}\}, \\ &\lambda \in (0, \infty] \cap v \in (0, \infty]. \end{aligned} \tag{30}$$

4.4. The backward marching scheme

Eq. (30) models the posterior distribution of the initial concentration field when T is known. In the problem of interest in this study, T is an unknown variable as well. Therefore, a backward marching scheme is used to reconstruct the entire history of the concentration field. The procedure is as follows:

1. Select a time step Δt and set $T = \Delta t$ and $t = t_{\text{current}}$ where t_{current} is the current time (arbitrary reference time).
2. Formulate a posterior in the form of Eq. (30) at $t = t_{\text{current}} - T$.
3. Compute the posterior mean estimate of concentration at t .
4. If the computed concentration reaches a value 1 at any location, exit the iteration. Else, set $T = T + \Delta t$ and return to step 2.

In this approach, the concentration prior to the measurement time is reconstructed backward in time until the releasing time is reached. Note that the sensitivity problems only need to be solved once in this approach (solve the sensitivity problem over a large time span and record concentration values at sensor locations at all time steps).

Computing integrals of the hierarchical posterior distribution Eq. (30) is not a trivial task. More importantly, one is often interested in the highest density region of the posterior distribution. Based upon these considerations, a Gibbs sampling based Markov chain Monte Carlo (MCMC) simulation method is used to compute the posterior mean estimate of concentration.

5. Numerical exploration of the posterior distribution

Monte Carlo simulation is used in this study to compute the posterior estimates of concentration. The estimator in Eq. (21) is approximated as

$$E_L(\theta|Y) = \frac{1}{L} \sum_{i=1}^L \theta^{(i)}, \tag{31}$$

where $\theta^{(i)}$'s ($i = 1:L$) are L samples randomly generated from the posterior distribution. By the law of large numbers, this sample mean converges to the true expectation as L goes to infinity. Hence, a large enough sample set will ensure the accuracy of the Monte Carlo approximation.

Key to Monte Carlo simulation is how to generate a reasonably large sample set from the high dimensional distribution Eq. (30). Markov chain Monte Carlo (MCMC) simulation algorithms generate samples from the posterior distribution while spending most of the sampling steps in the highest density regions of the posterior state space.

The essence of MCMC algorithms is to explore the state space of a random parameter using the Markov chain mechanism [6]. One of the key advantages of using MCMC is that one can draw samples even if $p(x)$ can only be evaluated up to the normalizing constant. The MCMC sampler designed for exploring Eq. (30) is based on the basic MCMC algorithm, the Metropolis–Hastings algorithm and the Gibbs algorithm. The pseudo-code is the following:

1. Initialize $\theta^{(0)}$, $\lambda^{(0)}$ and $v^{(0)}$
2. For $i = 0: N_{mcmc} - 1$
 - sample $\theta_1^{(i+1)} \sim p(\theta_1|\theta_{-1}^{(i+1)}, \lambda^{(i)}, v^{(i)}, Y)$
 - sample $\theta_2^{(i+1)} \sim p(\theta_2|\theta_{-2}^{(i+1)}, \lambda^{(i)}, v^{(i)}, Y)$
 - \vdots
 - sample $\theta_m^{(i+1)} \sim p(\theta_m|\theta_{-m}^{(i+1)}, \lambda^{(i)}, v^{(i)}, Y)$.
 - sample $u \sim U(0, 1)$
 - sample $\lambda^{(*)} \sim q_{\lambda}(\lambda^{(*)}|\lambda^{(i)})$
 - if $u < A(\lambda^{(*)}, \lambda^{(i)})$
 $\lambda^{(i+1)} = \lambda^{(*)}$
 - else
 $\lambda^{(i+1)} = \lambda^{(i)}$
 - sample $u \sim U(0, 1)$
 - sample $v^{(*)} \sim q_v(v^{(*)}|v^{(i)})$
 - if $u < A(v^{(*)}, v^{(i)})$
 $v^{(i+1)} = v^{(*)}$
 - else
 $v^{(i+1)} = v^{(i)}$.

In the above algorithm, N_{mcmc} is the total number of sampling steps, and $\theta^{(i)}$, $\lambda^{(i)}$, and $v^{(i)}$ are the samples generated in the i th iteration for θ , λ , and v , respectively. Also, $\theta_j^{(i)}$ is the j th component of $\theta^{(i)}$. The notation $\theta_{-j}^{(i+1)}$ denotes a $m - 1$ dimensional vector $\{\theta_1^{(i+1)}, \dots, \theta_{j-1}^{(i+1)}, \theta_{j+1}^{(i+1)}, \dots, \theta_m^{(i+1)}\}$. Also, $p(\cdot|\theta_{-j}^{(i+1)}, \lambda^{(i)}, v^{(i)})$ is the full conditional distribution of θ_j in the i th iteration and u is a random number generated from the standard uniform distribution $U(0, 1)$. Finally,

$q_{\lambda}(\cdot|\lambda^{(i)})$ and $q_v(\cdot|v^{(i)})$ are the proposal distributions for λ and v in the i th iteration, respectively.

This algorithm updates one component of θ at each sampling step. The proposal distribution for each component θ_j is its full conditional distribution, which is the probability distribution of θ_j conditional on all other components. This full conditional distribution is derived as follows:

$$p(\theta_j|\theta_{-j}, \lambda, v, Y) \sim N(\mu_j, \sigma_j^2), \tag{32}$$

$$\mu_j = \frac{b_j}{2a_j}, \quad \sigma_j = \sqrt{\frac{1}{a_j}}, \tag{33}$$

$$a_j = \sum_{s=1}^N \frac{H_{sj}^2}{v} + \lambda W_{jj}, \quad b_j = 2 \sum_{s=1}^N \frac{\mu_s H_{sj}}{v} - \lambda \mu_p, \tag{34}$$

$$\mu_s = Y_s - \sum_{t \neq j} H_{st} \theta_t, \quad \mu_p = \sum_{i \neq j} W_{ij} \theta_i + \sum_{k \neq j} W_{jk} \theta_k. \tag{35}$$

The acceptance probability of every Gibbs sample is 1; hence all the samples of θ generated in this way are accepted.

The second part of this algorithm uses an MH sampler to update λ and v . The proposal distributions used to generate new samples of λ and v are both normal distributions as follows:

$$q_{\lambda}(\lambda^{(*)}|\lambda^{(i)}) \sim N(\lambda^{(i)}, \sigma_{\lambda}^2) \tag{36}$$

and

$$q_v(v^{(*)}|v^{(i)}) \sim N(v^{(i)}, \sigma_v^2). \tag{37}$$

There are two notes to this sampling process. First, the physical limits of θ are 0 and 1. However, if such limits are posed to the sampling process (i.e. rejecting negative samples and samples greater than 1) the posterior mean estimate can never reach the limits, which causes biasedness. Therefore, in this study, no constraint is applied to the sampling process. By doing this, some physically unfeasible samples are generated, but the posterior mean estimates are feasible. Second, design of the proposal distributions $q_{\lambda}(\lambda^{(*)}|\lambda^{(i)})$ and $q_v(v^{(*)}|v^{(i)})$ must ensure that the effective regularization parameter ($\frac{1}{2} \lambda \sigma^2$) is not too large.

Once the designed Markov chain converges, the samples recorded thereafter are from the target posterior distribution Eq. (30). The posterior mean estimates can then be computed using these samples.

6. Numerical examples

In this section, the above introduced methodology is demonstrated via three numerical examples. Without loss of generality, the examples are studied in dimensionless form.

6.1. Example 1: 1D advection–dispersion in a homogeneous porous medium

The first example is a one-dimensional problem adopted from [1]. Inside the spatial domain $[0, 28]$, Eq. (1) holds

with unit constant velocity, porosity and dispersion coefficient ($u = 1.0, \phi = 1.0, D = 1.0$). The concentration values at $x = 0$ and $x = 28$ are kept at zero. The initial concentration is a rectangular pulse:

$$c_0(x) = \begin{cases} 1, & 13.5 \leq x \leq 14.5, \\ 0, & 0 \leq x \leq 13.5, \quad 14.5 \leq x \leq 28. \end{cases} \quad (38)$$

The concentration data are collected at $t = 2.0$, while the objective is to estimate the concentration at $t = 1.1$ and $t = 1.9$.

Following [1], the direct problem is solved on a grid with 112 elements with time step 0.02. The true and estimated concentration profiles at $t = 1.1$ and 1.9 are plotted in Figs. 1 and 2, respectively. It is noticed that the estimate at $t = 1.1$ is slightly better than the estimate at $t = 1.9$, which is not expected since the data are collected at $t = 2.0$. This is due to the fact that the posterior mean estimate is

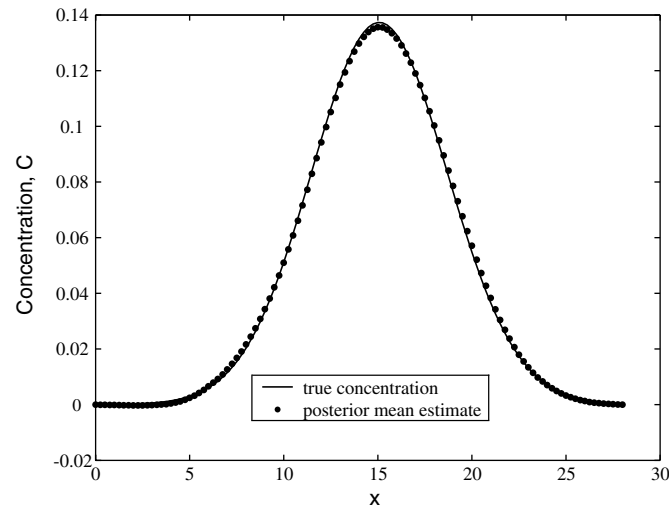


Fig. 1. True and posterior mean estimate of concentration at $t = 1.1$.

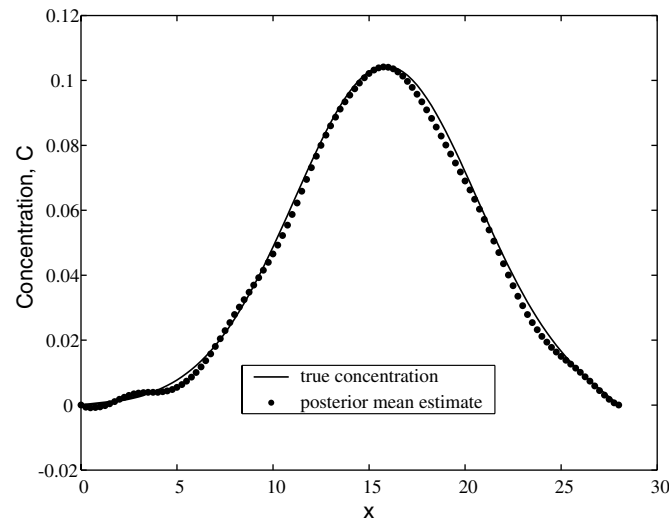


Fig. 2. True and posterior mean estimate of concentration at $t = 1.9$.

affected by the sample values of λ . Since the samples at these two time points were drawn from separate chains, the difference in sample values of λ causes the larger estimate error at $t = 1.9$. However, these two estimates are both rather accurate. This example is different from the one in [1] in that (i) the concentration data are assumed to be measured at 27 locations instead of at all element nodes and (ii) random noise with magnitude of 5% of the true concentration is added to the data. Still the estimates are quite accurate. The posterior density of the structure parameter λ in obtaining concentration estimate at $t = 1.1$ is plotted in Fig. 3 (Gamma distribution).

6.2. Example 2: 2D concentration reconstruction

In the following examples, we simulate a quarter area of the classical 5-spot problem. A schematic of the problem is shown in Fig. 4. Inside the square domain (unit side length), Eqs. (1) through (8) hold. The injection and production wells are located at $(0,0)$ and $(1,1)$, respectively, both having volume flux rate q (varying in different examples).

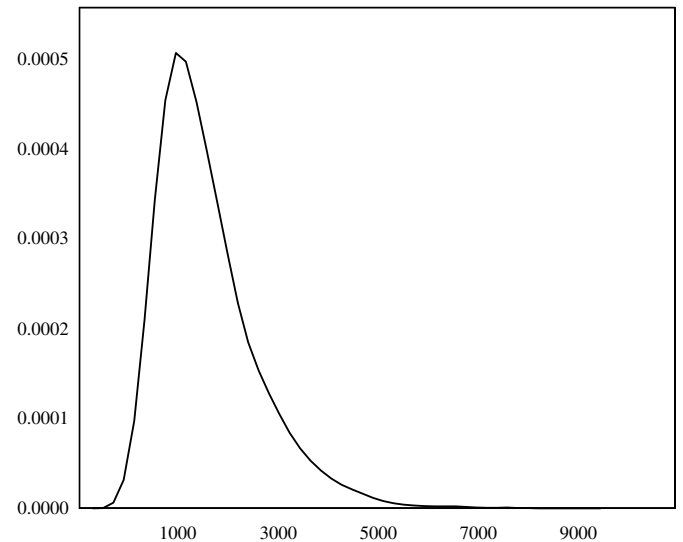


Fig. 3. Posterior density of structure parameter λ in obtaining concentration estimate at $t = 1.1$.

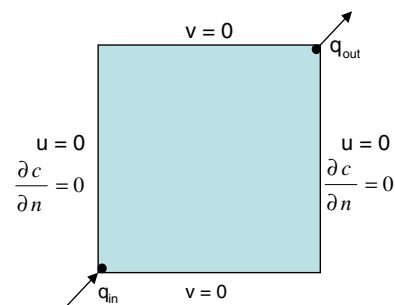


Fig. 4. Schematic of Example 2.

The actual initial concentration used to generate the simulation data has a normal distribution peaked at

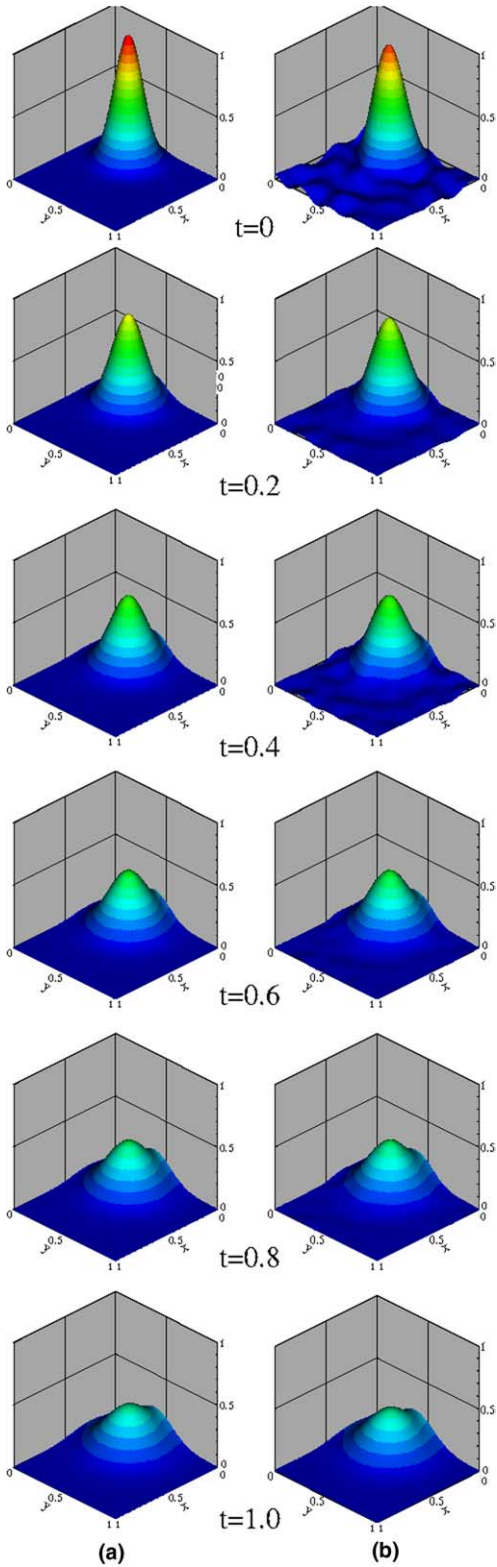


Fig. 5. Reconstruction of the history of contaminant concentration for diffusion-dominated transport in a homogeneous porous medium: (a) the true concentrations at different past time steps and (b) the reconstructed concentrations.

(0.375,0.75) with standard deviation 0.1. The direct problem is solved on a 128×128 finite element grid with a time step 0.02. Random measurement errors are simulated from a normal distribution with zero mean and standard deviation 0.005 (5–15% of the maximum recorded data in examples) in the homogeneous cases and 0.002 in the heterogeneous case. The simulation concentration data are generated by adding random measurement errors to the direct simulation solution at sensor locations. The data are used to recover the contaminant concentration history on a 64×64 grid. The number of sensors used varies in the following examples.

6.2.1. Case I: Diffusion-dominated transport in homogeneous porous media

In the first two-dimensional example, a diffusion-dominated mode is considered by setting a very small value ($q = 0.001$) for the well flux rate. The permeability and viscosity are both taken as 1.0. To ensure that molecular diffusion is the dominant transport mechanism, $\phi, \alpha_m, \alpha_1, \alpha_t$ take values 0.1, 0.1, 0.01 and 0.001, respectively. The concentration data are measured at $t = T = 1.0$ in this case. The sensors are evenly distributed on nodes of an 8×8 grid.

The true concentration profiles and corresponding reconstructed concentration profiles (posterior mean estimates) at different time points are plotted in Fig. 5. The time indices are obtained by setting $t_{\text{current}} = 1.0$. Since the concentration data are only collected at sparse sites at $t = 1.0$, it is of interest to reconstruct the entire concentration field at this time. This is performed here by solving the direct problem from $t = 0$ to $t = 1.0$ using the reconstructed concentration at $t = 0$ as the initial condition. The same backward step size is used as in the direct simulation ($\Delta t = 0.02$). It is seen that the estimated concentration profiles are rather close to the true concentration. The peak concentration value in posterior mean estimate at $t = 0$ is 0.9311, indicating that in this case, the backward marching procedure will be continued even after the true initial releasing time.

It has also been observed in the study that the posterior mean estimates at time points close to the measurement

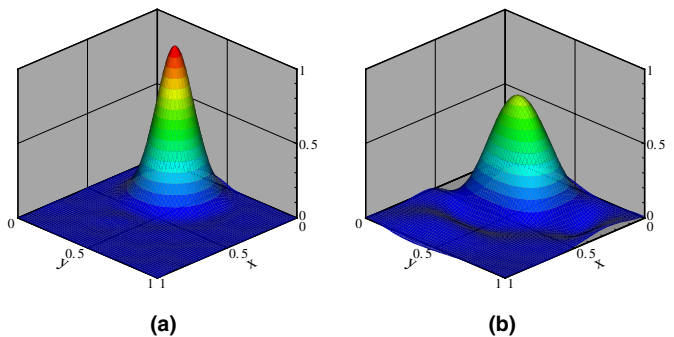


Fig. 6. Reconstruction of the concentration for diffusion-dominated transport in a homogeneous porous medium at $t = 0$: (a) data are collected at 9×9 sensor locations at $t = 0.2$ and (b) data are collected at 5×5 sensor locations at $t = 1.0$.

time are the most accurate. In Fig. 6(a), the concentration estimate at $t = 0$ (the true releasing time) using data mea-

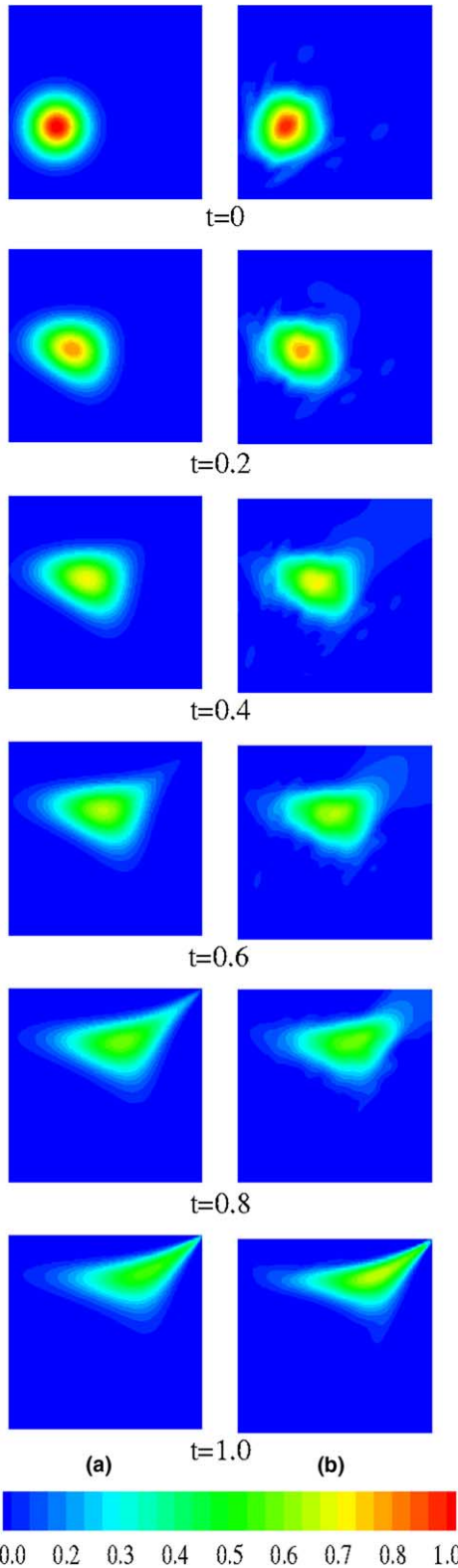


Fig. 7. Reconstruction of the history of pollutant concentration for advection-dominated transport in a homogeneous porous medium: (a) shows the true concentrations at different past time steps and (b) shows the reconstructions.

sured at $t = 0.2$ is plotted. The peak value in this case is 0.9916. The estimation of releasing time is very accurate in this case.

To test if the number of sensors can be further reduced, the above estimation is repeated using data at $t = 1.0$ from a 5×5 sensor network. The posterior mean estimate of the concentration at $t = 0$ (true releasing time) is plotted in Fig. 6(b). The peak value is only 0.8 in this case. Therefore, although the peak location and initial concentration profile can be identified in this case, the estimation of releasing time is not acceptable.

6.2.2. Case II: Advection-dominated transport in homogeneous porous media

In the second numerical experiment, we reconsider the earlier example by changing the following parameters: $q = 0.04$, $\alpha_m = 0$, $\alpha_l = 0.04$ and $\alpha_t = 0.004$. Convection and dispersion are the main mechanisms of contaminant propagation in this case. Fig. 7 shows the true concentration profiles and posterior mean estimates at different time points using data at $t = T = 1.0$. In this example, the data are measured using a 16×16 sensor network.

In Fig. 8, the estimated profile was generated using the data collected from a 9×9 sensor network. It is seen that

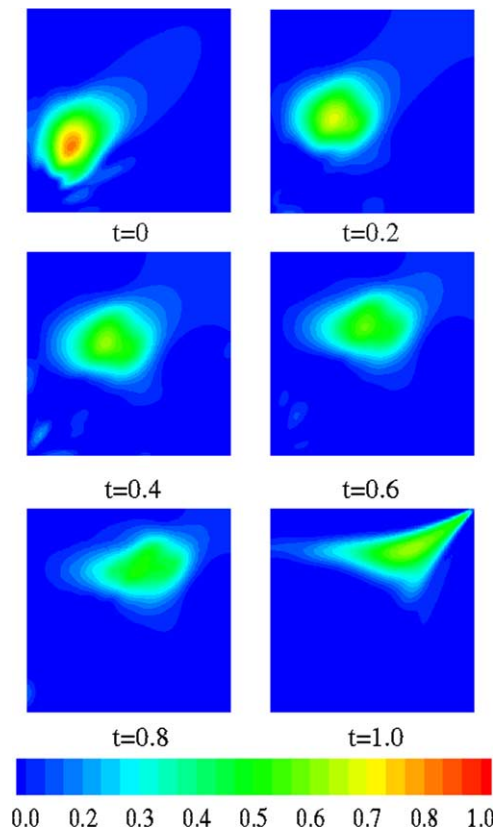


Fig. 8. Reconstruction of the contamination history for advection-dominated transport in a homogeneous porous medium when data are collected at 9×9 sensor locations at $t = 1.0$.

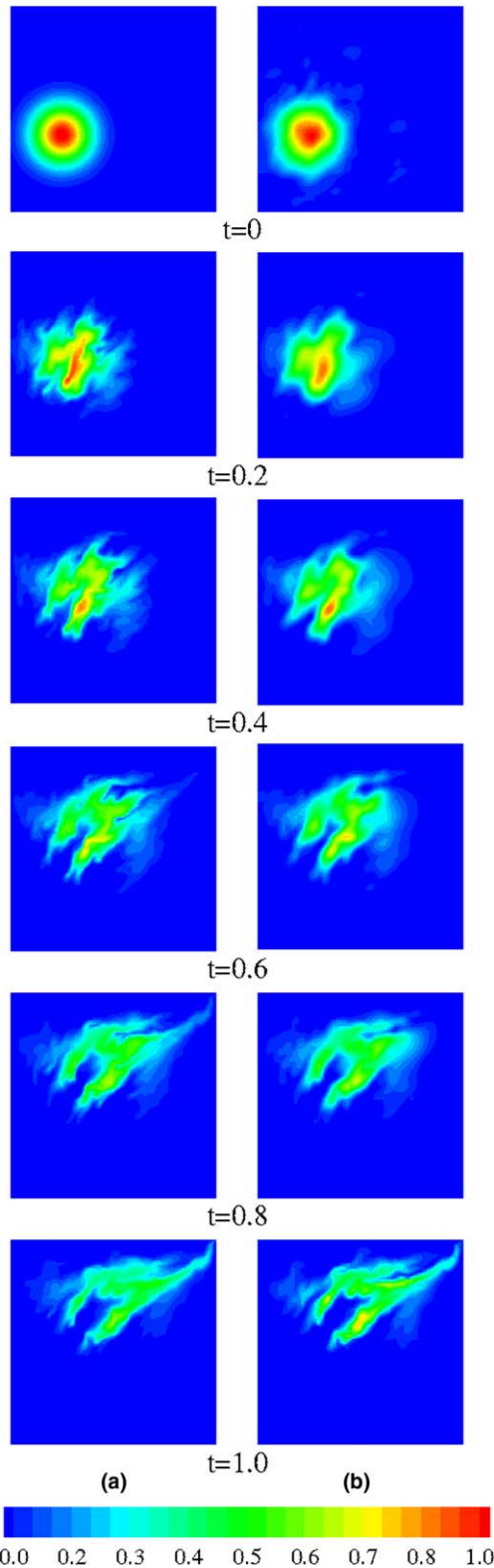


Fig. 9. Reconstruction of the history of pollutant concentration for advection–dispersion transport in a heterogeneous porous medium (data are collected at 32×32 grid): (a) the true concentrations at different past time steps and (b) the computed reconstructions.

more fluctuations exist in the estimates. However, the peak location and profile of concentration can still be resolved.

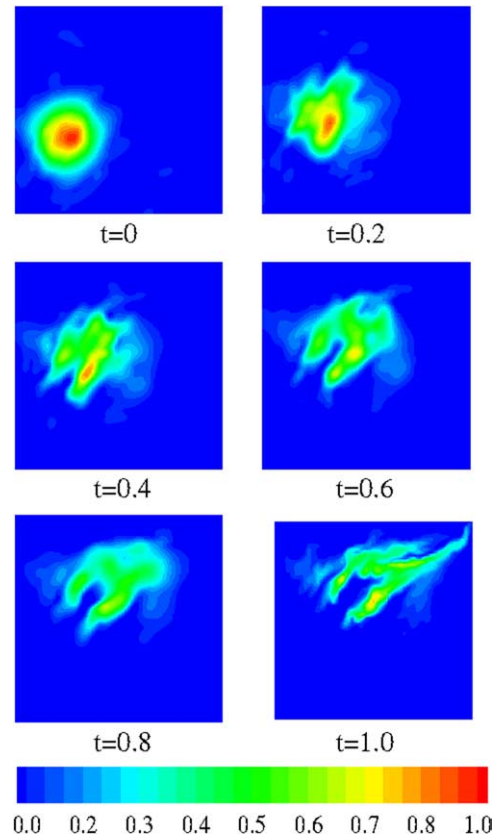


Fig. 10. Reconstruction of the history of pollutant concentration for advection–dispersion transport in a heterogeneous porous medium (data are collected on a 16×16 grid).

6.2.3. Case III: Advection–dispersion in heterogeneous porous media

In this example, we extend our earlier studies to heterogeneous porous media. All the quantities remain the same as in Example 2 in Section 6.2.2 except the permeability, which in this case is generated randomly from a joint log-normal distribution on a 32×32 finite lattice. The permeability mean at each site is 1.0 and the standard deviation of log permeability is 1.5. An uncorrelated permeability field is assumed in this case. The largest permeability and smallest permeability values in this example differ by the magnitude of 10^5 .

Fig. 9 shows the true concentration profiles and posterior mean estimates at different time points using data at $t = T = 1.0$. In this example, the data are measured from a 32×32 sensor network. The estimates obtained using data from a 16×16 sensor network are also presented in Fig. 10. It is observed that the estimates using less sensor data are comparable to the estimates in Fig. 9. Considering the heterogeneity and uniformly distributed sensor network, estimates in Fig. 10 are quite impressive.

7. Conclusions and discussion

A hierarchical Bayesian computation method is presented to solve the contaminant history reconstruction

problem in porous media. The regularity of the solution to this inverse problem is enforced by a pair-wise Markov random field model. Complete mathematical models are used in obtaining the direct simulation results of contaminant propagation in porous media flow. The attributes of the method are demonstrated via numerical examples in both homogeneous and heterogeneous porous media flows.

The current computational method successfully estimates instantaneously released contamination source in mixed fluids flow with constant viscosity. When the mobility ratio of contaminant to resident fluid (water) deviates largely from unity, a more complicated model is required to simulate the direct physical process. Furthermore, the estimation of continuously releasing contamination source is not addressed here and it is a problem of current research interest.

Acknowledgements

The work presented here was conducted using the resources of the Cornell Theory Center, which receives funding from Cornell University, New York State, federal agencies, and corporate partners.

References

- [1] J. Atmadja, A.C. Bagtzoglou, Pollution source identification in heterogeneous porous media, *Water Resour. Res.* 37 (2001) 2113–2125.
- [2] A.M. Michalak, P.K. Kitanidis, Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling, *Water Resour. Res.* 40 (2004) W08302.
- [3] M.F. Snodgrass, P.K. Kitanidis, A geostatistical approach to contaminant source identification, *Water Resour. Res.* 33 (1997) 537–546.
- [4] A.M. Michalak, P.K. Kitanidis, A method for enforcing parameter nonnegativity in Bayesian inverse problems with an application to contaminant source identification, *Water Resour. Res.* 39 (2003) 1033–1046.
- [5] A.M. Michalak, P.K. Kitanidis, Application of geostatistical inverse modeling to contaminant source identification at Dover AFB, Delaware, *J. Hydraul. Res.* 42 (2004) 9–18.
- [6] C. Andrieu, N.D. Freitas, A. Doucet, M.I. Gordan, An introduction to MCMC for machine learning, *Mach. Learn.* 50 (2003) 5–43.
- [7] J. Besag, P. Green, D. Higdon, K. Mengersen, Bayesian computation and stochastic systems, *Statist. Sci.* 10 (1995) 3–41.
- [8] D. Higdon, H. Lee, Z. Bi, A Bayesian approach to characterizing uncertainty in inverse problems using coarse and fine-scale information, *IEEE Trans. Signal Process.* 50 (2002) 389–399.
- [9] J. Besag, P.J. Green, Spatial statistics and Bayesian computation, *J. Royal Statist. Soc. Ser. B, Methodological* 55 (1993) 25–37.
- [10] J. Wang, N. Zabarar, A Bayesian inference approach to the stochastic inverse heat conduction problem, *Int. J. Heat Mass Transfer* 47 (2004) 3927–3941.
- [11] J. Wang, N. Zabarar, Hierarchical Bayesian models for inverse problems in heat conduction, *Inverse Problems* 21 (2005) 183–206.
- [12] P. Congdon, *Bayesian Statistical Modelling*, John Wiley & Sons, New York, 2001.
- [13] S. Geman, D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *Trans. Pattern Anal. Mach. Intell.* 6 (1984) 721–741.
- [14] W.R. Gilks, S. Richardson, D.J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, Chapman & Hall Ltd., New York, 1996.
- [15] A.F.D. Loula, F.A. Rochinha, M.A. Murad, Higher-order gradient post-processing for second-order elliptic problems, *Comput. Meth. Appl. Mech. Eng.* 128 (1995) 361–381.
- [16] J.P. Kaipio, E. Somersalo, *Computational and Statistical Methods for Inverse Problems*, Springer-Verlag, New York, 2005.
- [17] H. Wang, D. Liang, R.E. Ewing, S.L. Lyons, G. Qin, An ELLAM–MFEM solution technique for compressible fluid flows in porous media with point sources and sinks, *J. Comput. Phys.* 159 (2000) 344–376.
- [18] S. Balay, K. Buschelman, V. Eijkhout, W.D. Gropp, D. Kaushik, M.G. Knepley, L.C. McInnes, B.F. Smith, H. Zhang, *PETSc Users Manual*, ANL-95/11—Revision 2.1.5, Argonne National Laboratory, 2004.
- [19] A.L.G.A. Coutinho, J.L.D. Alves, Parallel finite element simulation of miscible displacement in porous media, *SPE J.* 1 (1996) 487–500.
- [20] A.L.G.A. Coutinho, C.M. Dias, J.L.D. Alves, L. Landau, A.F.D. Loula, S.M.C. Malta, R.G.S. Castro, E.L.M. Garcia, Stabilized methods and post-processing techniques for miscible displacements, *Comput. Meth. Appl. Mech. Eng.* 193 (2004) 1421–1436.
- [21] R.G. Sanabria Castro, S.M.C. Malta, A.F.D. Loula, L. Landau, Numerical analysis of space–time finite element formulations for miscible displacements, *Comput. Geosci.* 5 (2001) 301–330.