Journal of Hydrology 577 (2019) 123911

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Technical Note

A new inverse method for contaminant source identification under unknown solute transport boundary conditions

Jianying Jiao[a,*], Ye Zhang[a], Liqiang Wang[b]

[a] Department of Geology & Geophysics, University of Wyoming, Laramie, WY 82071, USA
[b] Department of Computer Science, University of Central Florida, Orlando, FL 32816, USA

ARTICLE INFO

This manuscript was handled by Huaming Guo, Editor-in-Chief, with the assistance of Mingjie Chen, Associate Editor

Keywords:
Solute transport
Contaminant source identification
LAS inverse method
Source release history
Boundary conditions

ABSTRACT

A new solute transport inverse method is proposed for estimating plume trajectory and source release location under unknown solute transport boundary conditions in a steady-state, non-uniform groundwater flow field. Solute concentration is modeled by proposing a set of local approximation solutions (LAS) of transport that are discretized over the problem domain. At a given time step, the inverse method imposes continuities of concentration and total solute mass flux at a set of collocation points in the inversion grid, whereas the LAS are conditioned to measured breakthrough concentrations. By enforcing transport physics at selected points in space and time, the inverse problem becomes well-posed and a single system of inversion equations is assembled and solved with a parallel iterative solver. Unlike most of the inversion techniques that minimize a model-data mismatch objective function, the inverse method does not require the simulation of a forward transport model for optimization, thus both solute initial and boundary conditions can be recovered. Assuming dispersivity estimates are available, the method was demonstrated using synthetic breakthrough data from various sampling densities and designs, i.e., irregular versus uniformly spaced well networks. Different measurement errors and source release histories (e.g., uniform-in-time, single, and multiple pulses) were also investigated. Results suggest that for the source release histories tested, 1) inversion is stable under increasing measurement errors up to 5% of the maximum observed concentration; 2) accurate plume trajectory and source release location can be estimated from solute breakthrough concentrations; 3) inversion accuracy appears the most sensitive to sampling well density and its information content.

1. Introduction

Worldwide, groundwater contamination from point- and non-point-sources is widespread. In the U.S., cleanup of over 300,000 soil and groundwater sites is projected to cost $200 billion by 2033 (National Research Council, 2013). Identification of pollutant pathways and sources is of particular importance for remediation design and the long-term planning and management of contaminated sites. Various techniques have been proposed for identifying contaminant pathways and sources and they can be divided into two general groups: (1) solute transport is solved with a forward- or backward-in-time numerical model with which plume trajectory is recovered by optimizing a model-date mismatch or objective function (Chadalavada et al., 2012; Mirghani et al., 2012; Yeh et al., 2016; Michalak and Kitanidis, 2004; Prakash and Datta, 2013; Sun, 2007). Because transport is dispersive and irreversible, inverse modeling using reverse time is considered ill-posed (Skaggs and Kabala, 1994). On the other hand, when a forward model is used, solute transport is simulated with a set of assumed initial and boundary conditions (BC) in order to provide simulated measurements to match with observations. For aquifers where hydraulic conductivity (K) is heterogeneous, various techniques have been extended to invert transport in non-uniform flows. To account for uncertainty in inversion, geostatistical methods have also been developed. (2) Solute transport is inverted using analytical solutions and regression techniques (Alapati and Kabala, 2000; Bagtzoglou and Atmadja, 2005; Sun et al., 2006; Woodbury and Ulrych, 1996; Bagtzoglou, 2003; Neupauer et al., 2000). To develop such solutions to describe plume migration, flow fields are usually considered uniform. Importantly, both groups of techniques assume that solute transport BC are known a priori in order to develop numerical and analytical solutions for optimization. At aquifer sites with limited or incomplete BC information, application of these techniques can lead to non-unique inverse solutions (Ayvaz, 2016).

To identify contaminant trajectory and source in a non-uniform flow

* Corresponding author.
E-mail address: jjiao1@uwyo.edu (J. Jiao).

https://doi.org/10.1016/j.jhydrol.2019.123911
Received 22 February 2019; Received in revised form 22 May 2019; Accepted 1 July 2019
Available online 02 July 2019

field, this study proposes a new inverse method by developing local approximate solutions (LAS) of concentration and by conditioning them to solute breakthrough measurements. Key difference between this new approach and the majority of existing inversion techniques is that standard regression techniques are not employed. Thus, a solute transport forward model does not need to be built and simulated in order to optimize a model-data mismatch, and as a result, non-uniqueness in inversion that is due to an incorrect assumption of transport BC is eliminated. (The same issue in groundwater flow inversion exists, which has been demonstrated in Irsa and Zhang, 2012). This article explains the new transport inverse method for estimating the state variable (i.e., concentration and its spatial-temporal evolution) assuming appropriate dispersivity data are available as prior information for inversion. The method was verified using synthetic concentration breakthroughs, with or without measurement errors, that were sampled from a "true" model that simulated several source release histories at a given source location. Note that the true model can be an analytical or numerical solution of the solute transport equation subject to a set of known initial and BC, or it can be a physical system created in a laboratory. In this article, because transport in non-uniform flows is of interest, a numerical model was adopted as the true model to test the inverse method.

Performance of the new inverse method was evaluated using various sampling densities and designs, i.e., irregular versus uniform well networks. The irregular network was based on monitoring wells from a shallow unconfined aquifer in the Texas High Plains where water level and core hydrogeological measurements were available. This aquifer, which overlies 4 Texas counties (Parmer, Castro, Bailey, Lamb), extends to a depth ~30 m and consists of unconsolidated coarse sand, gravel, minor clays, and carbonate minerals (Seni, 1980). From the numerous lithology logs and limited core measurements, the aquifer is moderately heterogeneous where K varies over ~3 orders of magnitude (http://www.twdb.texas.gov/groundwater/data/index.asp). Water level data came from USGS National Water Information System (https://waterdata.usgs.gov/nwis/nwis) and have been quality-checked by USGS with an approval status of "processing and review completed". From the water levels, overall groundwater flow direction is from west to east.

In the reminder of this article, the true model and the inverse method are described first, followed by a section describing application of the new method to plume recovery for different source release histories, e.g., uniform-in-time, single, and multiple pulses. Breakthrough data with varied measurement quality, density, and distribution were tested. We then summarize the outcomes while pointing out future research directions.

2. Method

A new inverse method is proposed for identifying solute plume history and source release location under unknown solute transport BC. The solute is assumed to be dilute, non-reactive, and is migrating in a steady state non-uniform flow field. Variable density coupling with flow is ignored. Note that groundwater flow and concentration data can be assimilated using a sequential inversion approach: (1) given hydraulic head and local K measurements, flow inversion is performed to recover a steady-state non-uniform flow field of an unconfined aquifer (Jiao and Zhang, 2014a, 2015a,b). (2) Given the flow field, solute plume history and source release location are identified by conditioning the estimated plume trajectory to observed breakthrough curves (BTC) at wells. For the above steps described, both BC for flow inversion and initial and BC for transport inversion are unknown. Two separate grids can be used based on measurement locations, e.g., one for flow inversion using wells with hydraulic data and one for transport inversion using wells with BTC data. As the flow inversion step has been previously described, this work focuses on plume identification within the estimated flow field. Below, equation describing the solute transport true model is presented first, followed by a description of the inverse method. Groundwater is assumed to have negligible vertical flow, thus transport inversion is performed in 2D on the horizontal plane. Moreover, the physics governing solute transport is assumed known and is described by the advection-dispersion equation (ADE).

To verify the inverse method, numerical transport experiments ("true model") were performed solving the ADE subject to different source release scenarios to provide synthetic measurements for inversion:

$$\frac{\partial c(x,y,t)}{\partial t} = \frac{\partial}{\partial x}\left((\alpha_L |\boldsymbol{v}| + D_s)\frac{\partial c(x,y,t)}{\partial x}\right) + \frac{\partial}{\partial y}\left((\alpha_T |\boldsymbol{v}| + D_s)\frac{\partial c(x,y,t)}{\partial y}\right) - \left(\frac{\partial}{\partial x}(c(x,y,t)v_x(x,y)) + \frac{\partial}{\partial y}(c(x,y,t)v_y(x,y))\right) \quad on\ \Omega \tag{1}$$

$$c(x,y,0) = 0 \quad on\ \Omega \tag{2}$$

$$c(x,y,t) = c_0(x,y) \quad on\ \Gamma_1, \quad t > 0 \tag{3}$$

where Ω is computational domain; c is concentration in mass per unit pore water volume [M/L^3]; D_s is porous medium diffusion coefficient. $\boldsymbol{v} = [v_x(x,y), v_y(x,y)]$ is average linear velocity [L/T]; $\boldsymbol{v}$ is related to Darcy flux $\boldsymbol{q}$=$[q_x(x,y), q_y(x,y)]$ via effective porosity θ ($\boldsymbol{v} = q/\theta$) which is assumed a known constant; α_L and α_T are local longitudinal and transverse dispersivities, respectively, assumed known as priori information. We ignore spatial variation of α_L and α_T because the site K field resolved by flow inversion is moderately heterogeneous. Finally, $\Gamma_1$is Dirichlet BC where zero concentration is assigned at the inflow boundary. Zero mass flux is assigned to the top and bottom while outflow BC are imposed on the remaining boundary so solute can migrate out of the domain.

For the same computational domain Ω, transport inversion enforces 3 sets of constraints: (1) global continuity of concentration and total solute flux mass at a set of collocation points at a given (discretized) time; (2) local conditioning of the inverse solutions, of the given time, to observed breakthrough concentrations at the same time; (3) equation constraints that enforces the transport physics at selected points in space and time. The first set of constraints is written as:

$$w^n(p_\gamma)\int R_{c^n}(p_\gamma)\delta^n(p_\gamma)d\Gamma_m = 0, \quad m = 1, \cdots, Y; \quad \gamma = 1, \cdots, V \tag{4}$$

$$w^n(p_\gamma)\int R_{\boldsymbol{J}^n}(p_\gamma)\delta^n(p_\gamma)d\Gamma_m = 0, \quad m = 1, \cdots, Y; \quad \gamma = 1, \cdots, V \tag{5}$$

Where n denotes a discretized inversion time, p_γ denotes a collocation point on mth cell interface (Γ_m) in the inversion grid, γ is the number of collocation points on Γ_m, and $R_{\boldsymbol{J}^n}(p_\gamma)$ are residuals of solute concentration and total solute mass flux at p_γ on Γ_m at the nth time, respectively, Y is the number of cell interfaces in the grid, $\delta^n(p_\gamma)$ is Dirac delta function at the nth time, and $w^n(p_\gamma)$ is weighting function which samples the residuals at p_γ on Γ_m at the nth time. Details on how the weight is implemented can be found in Zhang et al. (2014). The residuals can be expanded as:

$$R_{c^n}(p_\gamma) = c_i^n(p_\gamma) - c_k^n(p_\gamma) \tag{6}$$

$$R_{\boldsymbol{J}^n}(p_\gamma) = \boldsymbol{J}_i^n(p_\gamma) - \boldsymbol{J}_k^n(p_\gamma) \tag{7}$$

$$\boldsymbol{J}^n = \boldsymbol{v}c^n - \boldsymbol{D}\nabla c^n \tag{8}$$

where c is a proposed LAS that varies with space and time: c_i^n and c_k^n denote concentrations of cells i and k adjacent to Γ_m at the nth time; $\boldsymbol{D}$ is the dispersion tensor; $\boldsymbol{J}^n$ is a proposed LAS for total solute mass flux: $\boldsymbol{J}^n$depends on c and ∇c thus is space and time dependent. For a two-dimensional problem, the total mass flux can be written as (Bear and Cheng, 2010): $\boldsymbol{J}^n = [J_x^n, J_y^n] = \left[v_x(x,y)c^n - (\alpha_L |\boldsymbol{v}| + D_s)\frac{\partial c^n}{\partial x}, v_y(x,y)c^n - (\alpha_T |\boldsymbol{v}| + D_s)\frac{\partial c^n}{\partial y}\right]$.

The second set of constraints is defined by conditioning c^n to measured concentrations:

$$w^n(p_a)(c^n(p_a) - c_{ob}^n(p_a)) = 0 \quad a = 1, \cdots, A; \tag{9}$$

where p_a is a measurement point; c_{ob}^n is observed concentration at p_a at

the nth time (A is total number of observations); and $w^n(p_a)$ is a weighting function at the nth time assigned to each observation equation to reflect the magnitude of measurement errors. $w^n(p_a)$ can be time dependent (i.e., becomes larger) if measurement quality improves with time.

The inverse method assumes that the governing transport equation is known, which provides a set of equation constraints for inversion:

$$w^n(p_e)R_e^n = \varepsilon \tag{10}$$

$$R_e^n = \left[\frac{\partial c^n}{\partial t} - \frac{\partial}{\partial x}\left((\alpha_L\,|\boldsymbol{v}| + D_s)\frac{\partial c^n}{\partial x}\right) - \frac{\partial}{\partial y}\left((\alpha_T\,|\boldsymbol{v}| + D_s)\frac{\partial c^n}{\partial y}\right) + \left(\frac{\partial}{\partial x}(c^n v_x) + \frac{\partial}{\partial y}(c^n v_y)\right)\right]\Bigg|_e$$

$$e = 1,\ \cdots, YV + A$$

where p_e include both collocation points and measurement locations (p_e can be time-dependent, which is not implemented here); R_e is the equation residual at p_e at the nth time. For a given time, Eq. (10) enforces a set of physical constraints on the LAS of c^n and $\boldsymbol{J}^n$ at p_e. The system of inverse equations is solved using an iterative solver, where ε is on the order of 10^{-5} or smaller.

In Eqs. (4)–(10), both LAS of c^n and $\boldsymbol{J}^n$ are approximate rather than exact solutions, which allows a flexible handling of parameters, initial, and BC, while Eq. (10) provides physics-based constraints on the approximations. In the solution domain, $v_x(x, y)$ and $v_y(x, y)$ are obtained from flow inversion and are deterministic coefficients in the LAS. For each inversion grid cell (Ω_e), polynomial functions were chosen as the LAS:

$$c^n(x, y, t) = a_1^n + a_2^n t_n + (a_3^n + a_4^n t_n)x + (a_5^n + a_6^n t_n)y + (a_7^n + a_8^n t_n)xy \tag{11}$$

where $a_i^n (i = 1, \cdots, 8)$ are unknown, time-dependent coefficients that together control the shape of the LAS of concentration. Following Eq. (11), total mass flux can be approximated as:

$$\begin{aligned} J_x^n(x, y, t) &= c^n v_x - (\alpha_L\,|\boldsymbol{v}| + D_s)\frac{\partial c^n}{\partial x} \\ &= (a_1^n + a_2^n t_n + (a_3^n + a_4^n t_n)x + (a_5^n + a_6^n t_n)y + (a_7^n + a_8^n t_n) \\ &\quad xy)v_x - (\alpha_L\,|\boldsymbol{v}| + D_s)((a_3^n + a_4^n t_n) + (a_7^n + a_8^n t_n)y) \end{aligned} \tag{12}$$

$$\begin{aligned} J_y^n(x, y, t) &= c^n v_y - (\alpha_T\,|\boldsymbol{v}| + D_s)\frac{\partial c^n}{\partial y} \\ &= (a_1^n + a_2^n t_n + (a_3^n + a_4^n t_n)x + (a_5^n + a_6^n t_n)y + (a_7^n + a_8^n t_n) \\ &\quad xy)v_y - (\alpha_T\,|\boldsymbol{v}| + D_s)((a_5^n + a_6^n t_n) + (a_7^n + a_8^n t_n)x) \end{aligned} \tag{13}$$

Following Eq. (11), R_e^n inEq. (10) is rewritten as:

$$\begin{aligned} &a_2^n + (a_4^n)x + (a_6^n)y + (a_8^n)xy + (v_x((a_3^n + a_4^n t_n) + (a_7^n + a_8^n t_n)y) + v_y \\ &\quad ((a_5^n + a_6^n t_n) + (a_7^n + a_8^n t_n)x)) = 0 \end{aligned} \tag{14}$$

For the discretized inversion spatial and temporal domains, Eqs. (4)–(10) can be assembled to form a single system of equations that can be solved using standard techniques. A scalable parallel linear solver based on the Message Passing Interface, Scalable Parallel LSQR (Huang et al., 2013), is implemented for its solution, for which the inversion coefficient matrix is decomposed to a kernel sub-matrix and a damping sub-matrix to reduce overlapping information. This decreases interprocessor communication and achieves computational efficiency (Wang, 2014). For a trial transport inversion problem, a serial solver and the parallel solver were both employed, yielding nearly identical results.

3. Results

3.1. Computational domain

A regional-scale, 2D computation domain (112 km E-W and 79 km N-S) was selected to reflect realistic monitoring well designs from an unconfined aquifer in Texas High Plains. Before transport inversion was performed, steady state flow was inverted to recover long-term water table and Darcy fluxes by conditioning to core K and water level data from 329 (active) wells. To obtain groundwater velocity, an average effective porosity from core measurements was used. Because field measurements are never exhaustive, uncertainty exists in the inverted flow field. However, it is treated here as deterministic, and is imported into the transport analysis for both the true/forward model and the inverse analyses. Specifically, based on observed concentrations, prior dispersivity estimates, and the groundwater velocity, solute concentration was inverted without the knowledge of both solute initial and boundary conditions. Moreover, based on site lithology data and mean groundwater velocity, an average Pe number of ~1.0 was calculated assuming the molecular diffusion coefficient of common dissolved hydrocarbons. Thus, for the modeled site, solute transport is advection-dominated (Perkins and Johnston, 1963; Batu, 2005) and diffusion is ignored in both the forward model and the inverse analysis.

3.2. Numerical true model

Four transport models were simulated in the domain, solving Eqs. (1)–(3) using MT3DMS [Zheng, 2010]. In each model, a tracer was released in upstream flow field (x = 19.1–19.9 km, y = 55.2–56.1 km) near the left boundary and migrated towards the right boundary. One model simulated continuous (constant-in-time) source release for a period of 1.8×10^6 days, which is also the total simulation time (Fig. 1a). For a shorter simulation time of 8×10^5 days, the second model simulated a short (10^4 day) release of the tracer at the same location (Fig. 1b). During the release time, a constant concentration of 10 mg/l is specified at the source location in both models. For the simulation time of 8×10^5 days, two additional release scenarios were simulated, with a Gaussian and multimodal release at the source, respectively (Fig. 1c,d). In the last example, 4 Gaussian distributions were superimposed at the source to create 4 peak concentrations over time. In all simulations, α_L and α_T, of 0.1 km and 0.03 km, respectively, were assigned to the grid cells (Gelhar et al., 1992). For the true model, MT3DMS implements the finite difference scheme ("FDM") for spatial and temporal discretization. Various discretizations were experimented to ensure the accuracy of transport simulations. The final grid has 1120 × 790 cells (Δx = Δy = 100 m) and Δt = 1 day. From the FDM, concentrations were sampled at hypothetical well locations and then imposed with unbiased measurement errors:

$$c^m = c^{FDM}(1 + \sigma \cdot rand) \tag{15}$$

where σ is absolute percent error and *rand* is a uniform r.v. from [−1, 1]. In inversion, the highest σ tested is 5% of maximum concentration in the FDM.

To assess the accuracy of inversion, a root mean square error (RMS) and a relative root mean square error (RES) of concentration were computed at each well location at a given sampling time (t):

$$RMS(t) = \sqrt{\frac{1}{N_{total}}\sum_{i=1}^{N_{total}}(c_{true}(x_i, y_i, t) - c(x_i, y_i, t))^2} \tag{16}$$

$$RES(t) = \frac{\sqrt{\sum_{i=1}^{N_{total}}(c_{true}(x_i, y_i, t) - c(x_i, y_i, t))^2}}{\sqrt{\sum_{i=1}^{N_{total}}(c_{true}(x_i, y_i, t))^2}} \tag{17}$$

where $c_{true}(x_i, y_i, t)$ is error-free measured concentration from the FDM, $c(x_i, y_i, t)$ is inverted concentration at the same location at the same

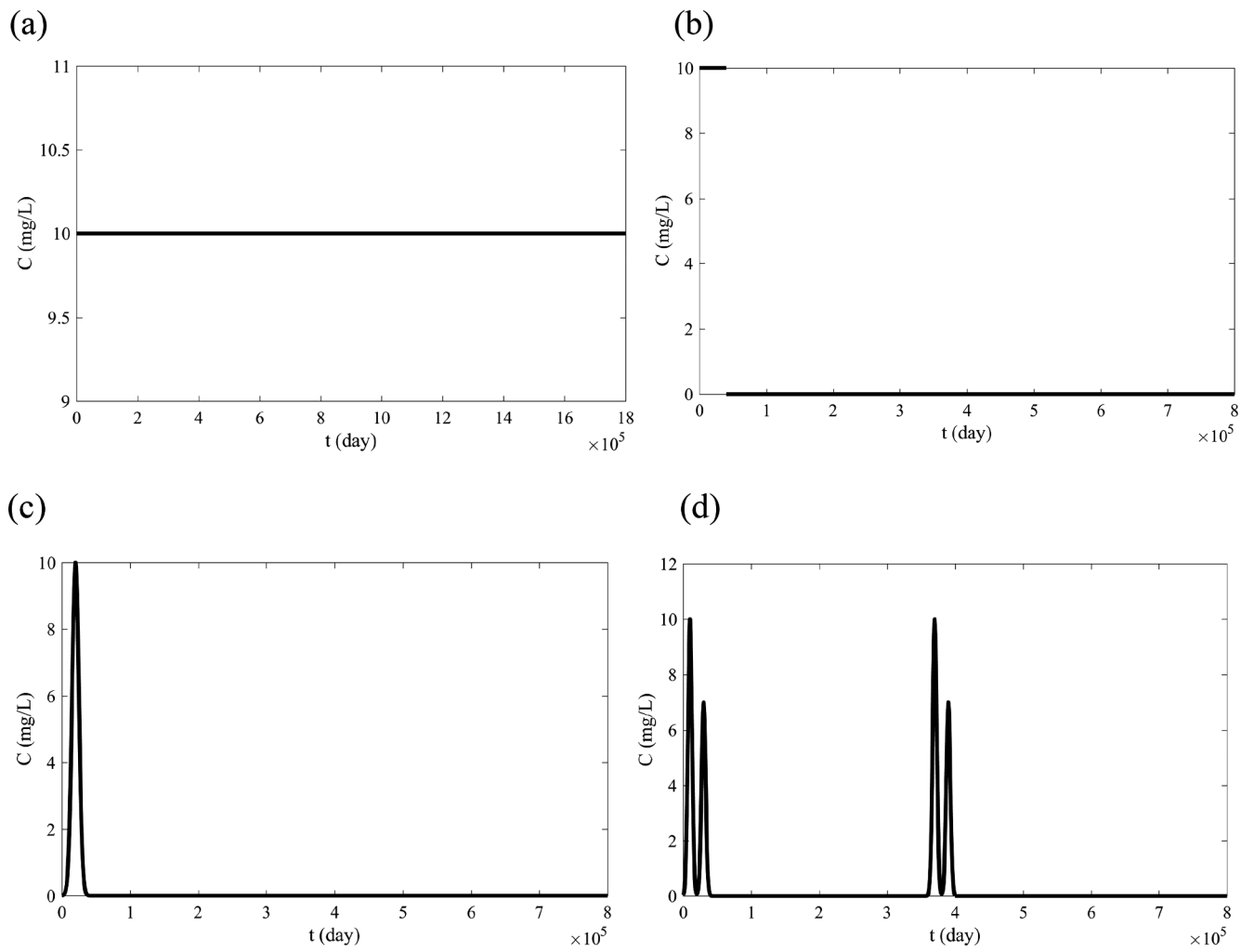

Fig. 1. Source release history: (a) Tests 1–7; (b) Test 8; (c) Test 9; (d) Test 10.

Table 1
Inversion tests for both continuous (Tests 1–7) and short-duration (Test 8–10) source release scenarios. The sampling wells either follows a gridded design (i.e., gridded) or a design taken from the Texas field site.

	Computational domain (km)	Inverse grid discretization	Number of observation wells	Longitude dispersivity (km)	Transverse dispersivity (km)	Absolute percent error in observed concentrations	RMS (t1)	RES (t1)
Test 1	112 × 79	56 × 40	2184 (gridded)	0.1	0.03	0%	6.5×10^{-3}	5.5×10^{-3}
						5%	1.2×10^{-1}	7.1×10^{-2}
Test 2	112 × 79	14 × 10	329	0.1	0.03	0%	1.2×10^{-1}	9.6×10^{-2}
						5%	1.4×10^{-1}	9.9×10^{-2}
Test 3	112 × 79	14 × 10	329	1	0.3	5%	3.7×10^{-1}	2.4×10^{-1}
Test 4	112 × 79	14 × 10	329	0.01	0.003	5%	6.9×10^{-2}	4.6×10^{-2}
Test 5	112 × 79	14 × 10	140 (gridded)	0.1	0.03	0%	2.7×10^{-2}	1.2×10^{-2}
						5%	4.2×10^{-1}	1.8×10^{-1}
Test 6	40 × 40	10 × 10	65	0.1	0.03	0%	1.6×10^{-2}	2.4×10^{-2}
						5%	2.3×10^{-2}	4.5×10^{-2}
Test 7	40 × 40	10 × 10	100 (gridded)	0.1	0.03	0%	1.6×10^{-2}	2.3×10^{-2}
						5%	1.5×10^{-2}	2.3×10^{-2}
Test 8	112 × 79	56 × 40	2184 (gridded)	0.1	0.03	0%	5.5×10^{-5}	4.0×10^{-2}
						5%	5.4×10^{-4}	4.7×10^{-1}
Test 9	112 × 79	56 × 40	2184 (gridded)	0.1	0.03	0%	1.4×10^{-6}	4.2×10^{-2}
						5%	1.8×10^{-5}	4.8×10^{-1}
Test 10	112 × 79	56 × 40	2184 (gridded)	0.1	0.03	0%	1.8×10^{-4}	2.1×10^{-2}
						5%	1.9×10^{-3}	4.0×10^{-1}

time, and N_{total} is total number of measurements at t. From each synthetic well in the FDM, solute BTC was sampled at two discrete times (t_1, t_2). The recovered plume was estimated at these times for which RMS and RES were computed.

3.3. Continuous source release

For the release history in Fig. 1a, 7 tests were carried out to evaluate the accuracy and stability of inversion under different σ, well density/

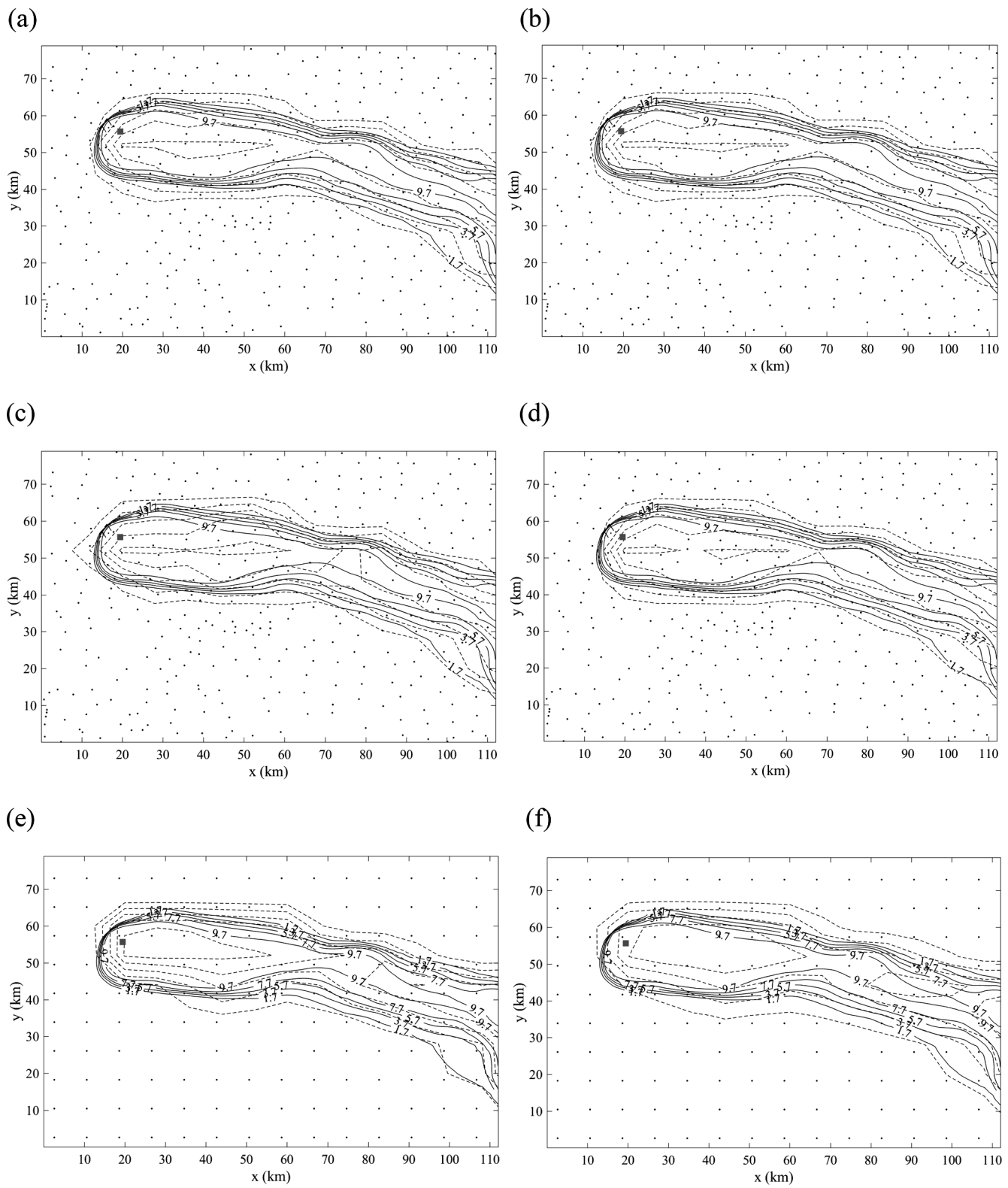

Fig. 2. Inverted concentration (dashed contours; mg/l) compared to true concentration simulated by the FDM (solid contours; mg/l) at $t_1 = 1.6 \times 10^6$ day. Shaded box indicates the source release location; dots indicate the location of sampling wells. (a) 0% error of Test 2; (b) 5% error of Test 2; (c) 5% error of Test 3; (d) 5% error of Test 4; (e) 0% error of Test 5; (f) 5% error of Test 5.

sampling pattern, and inversion domain size and discretization (Table 1). For most tests, σ was increased from 0% to 5%. Test 1 employs a large number of sampling wells (2184) distributed in a gridded pattern (average well spacing is 2 km). For this test, a relatively large inversion grid (56×40) was used, coarsened from the true model (1120×790) by ~20 times in each dimension. True α_L and α_T were assigned as prior information for inversion. Tests 2–4 use fewer (329) sampling wells following the pattern of the Texas aquifer. A coarsened inversion grid (14×10) was used. To test sensitivity of inversion to the assumed values of α_L and α_T, these were modified by 1 order of

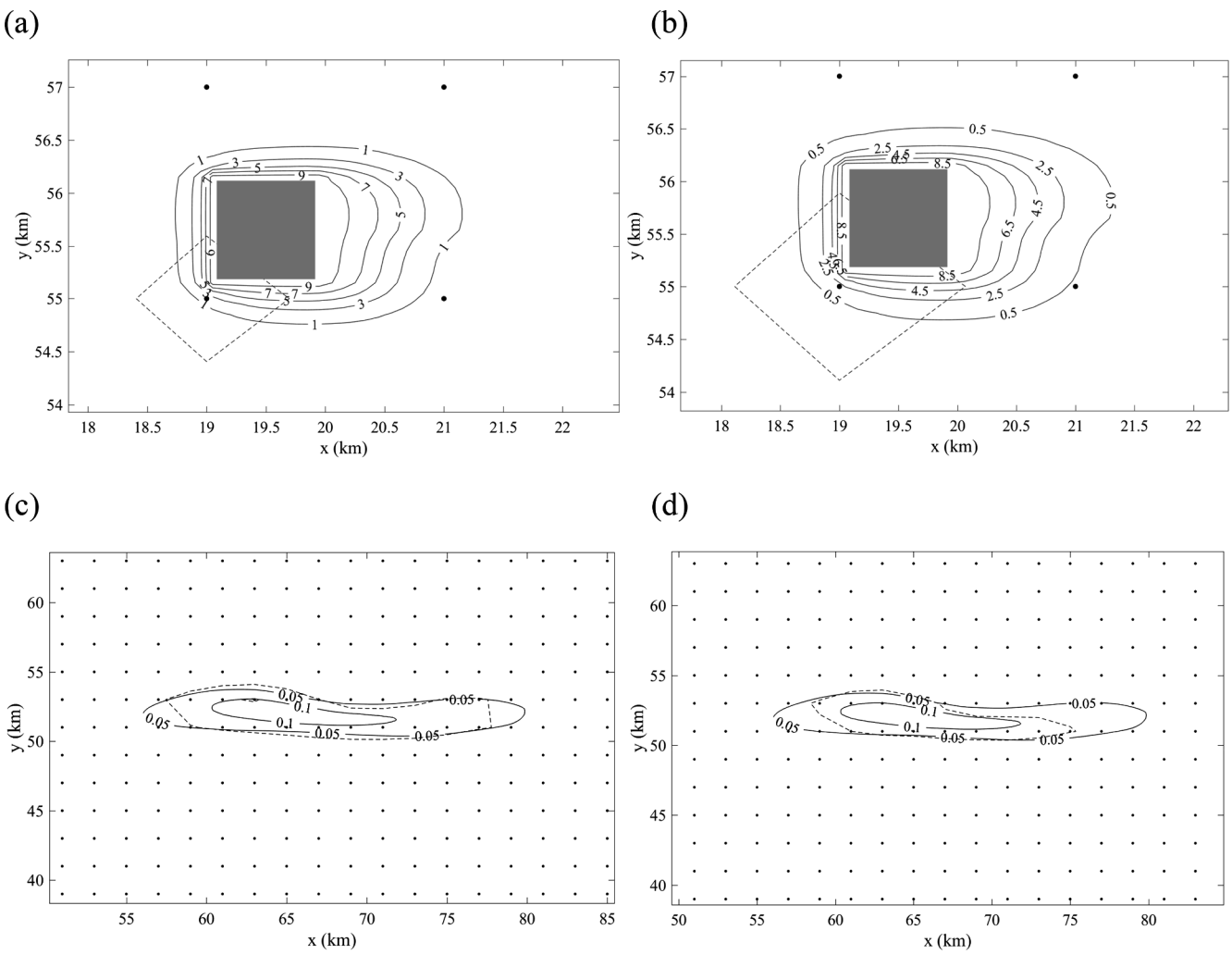

Fig. 3. Inverted concentration (dashed contours) of Test 8 compared to true concentration simulated by the FDM (solid contours) at $t_1 = 1.0 \times 10^4$ day (top row) and $t_2 = 8.0 \times 10^5$ day (bottom row). Shaded box indicates the source release location; dots indicate the location of sampling wells. (a) 0% error; (b) 5% error; (c) 0% error; (d) 5% error.

magnitude, introducing errors into inversion. Compared to Test 2, Test 5 further reduces sampling density although wells were placed in a gridded pattern. For Tests 1–5, observed concentrations were sampled at $t_1 = 1.6 \times 10^6$ and $t_2 = 1.8 \times 10^6$ days since release of the solute: for most wells, these times largely captured the *late-time* portion of the BTCs. In Tests 6–7, BTCs at $t_1 = 4.0 \times 10^4$ and $t_2 = 4.1 \times 10^4$ days were sampled from the FDM, which captured the *early-time* portion of the BTCs. Accordingly, a smaller domain closer to the source location was used (x = 5–45 km, y = 35–75 km) in inversion along with a smaller grid (10 × 10). Test 6 adopts the field network while Test 7 uses a gridded pattern. For all tests, RMS and RES of the recovered concentrations at t_1 are tabulated. At t_2, similarly valued error measures were computed (not shown).

Accuracy of plume recovery is evaluated by RMS and RES, and by comparing the recovered plume against that simulated by FDM (Fig. 2). For all test cases, inversion outcomes are stable and compare favorably with the true plume and are relatively insensitive to measurement error, inversion grid discretization, and dispersivities (prior information). In contrast, the number of wells used to condition inversion has the largest impact on accuracy. Test 1, with the largest number of observation wells, yields the lowest RMS and RES (for σ = 0, they vary from 6.5×10^{-3} and 5.5×10^{-3}). When sampling density is reduced, inversion becomes less accurate. Moreover, even though Test 5 has a lower density than Test 2, its RMS and RES are actually lower. This is attributed to the gridded sampling pattern with improved information content. To test this idea, two additional cases (not listed in Table 1) were created for the problem domain of Tests 1–5. One used an irregular network from the Texas aquifer with 1650 wells, including both active (329) and abandoned wells (1321). From the FDM, 529 of the 1650 wells were observed to provide positive BTCs. The other used gridded sampling, with a total of 609 wells for which only 191 wells have positive BTCs. For both error-free and increased errors in measured concentrations, plume recovery is more accurate for the gridded network. Irregular sampling likely suffers from clustering, whereby closely spaced wells with similar measurements exert undue influences on data conditioning during inversion. Further research will evaluate optimal weights for which de-clustering methods will be investigated (Li et al., 2018).

In Tests 6 and 7, the FDM plume was sampled at earlier times when the plume was smaller. Inversion adopts a smaller computational domain with correspondingly fewer wells that lie close to the source location. For Test 6, plume was sampled following the field network; Test 7 sampled the same plume in a gridded pattern. At $t_1 = 4.0 \times 10^4$ day, only 11 (Test 6) and 16 wells (Test 7) have positive BTCs while the remaining 54 (Test 6) and 84 wells (Test 7) have no BTC as they lie outside the plume. Inversion is fairly accurate and robust for both tests (not shown), although Test 7 (gridded network) yields a slightly better recovered plume.

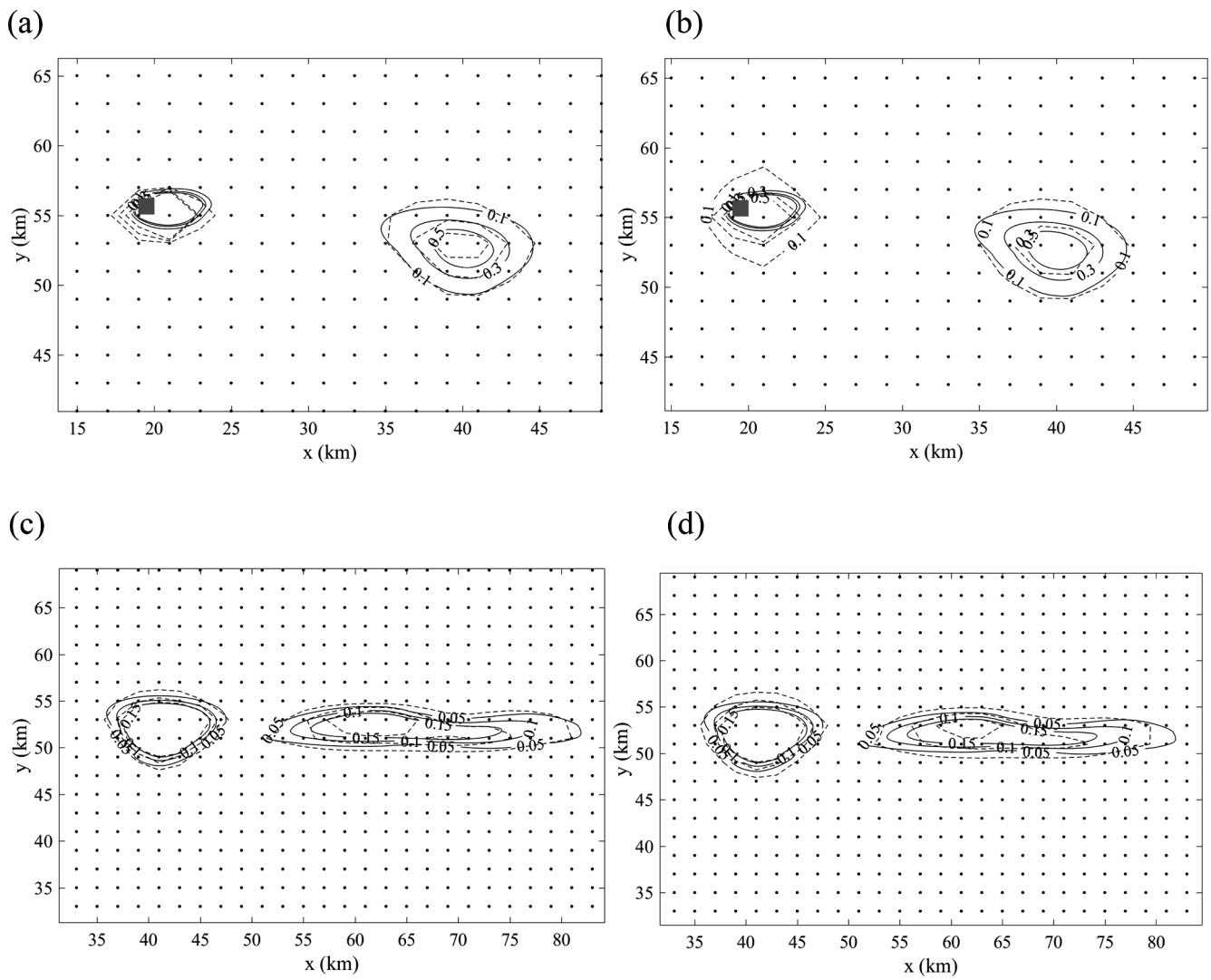

Fig. 4. Inverted concentration (dashed contours) of Test 10 compared to true concentration simulated by the FDM (solid contours) at $t_1 = 4.0 \times 10^5$ day (top row) and $t_2 = 8.0 \times 10^5$ day (bottom row). Shaded box indicates the source release location; dots indicate the location of sampling wells. (a) 0% error; (b) 5% error; (c) 0% error; (d) 5% error.

3.4. Pulse releases (uniform, Gaussian, multimodal)

For a pulse release scenario (Fig. 1b; Test 8), inversion was carried out using the gridded sampling of Test 1. Because less solute mass was released, plume extent was more limited and $t_1 = 1.0 \times 10^4$ days and $t_2 = 8.0 \times 10^5$ days were chosen as sampling times. Though 2184 observation wells were used in inversion, only a very small subset (4 at t_1 and 24 at t_2) has positive BTCs. Result was stable under increasing σ (Fig. 3), although accuracy is lower at t_1 when only 4 wells have BTCs. At t_1, which concurs with the end of tracer release, inversion has largely identified the release location despite the inaccurate plume shape. At t_2, the plume evolved to be larger and migrated further downstream. More wells have positive BTCs, and accordingly, the inverted plume was very accurate. In addition, using the sampling pattern of Test 2 (field network), the pulse release was inverted again. Because none of the wells have positive BTCs at t_1 and t_2, inversion cannot recover the plume, as expected.

In Tests 9 and 10, release history based on a single to multiple superimposed Gaussian functions (Fig. 1c,d) was simulated by the FDM and subsequently inverted using the gridded sampling of Test 1 (Table 1). While true dispersivities were given as prior information, σ was again increased. Both RMS and RES and plume recovery suggest that accurate inversion can be obtained for these cases. For Test 10, inversion was able to capture the bi-modal plume shape that had evolved in both early and late times (Fig. 4).

4. Conclusion

A new inverse method, based on approximating solute concentration and mass flux with local approximate solutions (LAS), is proposed for estimating plume trajectory and source release location under unknown solute transport BC. At a given inversion time, the method imposes concentration and flux continuities at a set of collocation points in space, while the LAS are conditioned to measured breakthrough concentrations. By enforcing transport physics at selected points in space and time, the inverse problem becomes well-posed. A single system of equations is assembled and solved with a parallel linear solver. Unlike the majority of existing techniques, plume trajectory and source location can be recovered efficiently because the new method does not require the repeated simulations of a forward transport model. For an unconfined aquifer with a moderately heterogeneous flow field, the new method was demonstrated by inverting synthetic concentration breakthroughs from well networks sampling a numerical true model. Different measurement errors and source release histories (e.g., uniform-in-time, single, and multiple pulses) were evaluated. Results suggest that for the source release histories tested, (1) inversion is

stable under increasing measurement errors up to 5% of the maximum observed concentrations; (2) accurate plume recovery and source release location can be attained from the BTC data; (3) inversion accuracy is the most sensitive to the sampling well density and its information content. For the cases investigated herein, the recovered plume also appears relatively insensitive to the assumed dispersivity values.

This work extends our previous research inverting transient groundwater flow under unknown fluid flow initial and BC (Jiao and Zhang, 2014a,b, 2015b) to contaminant source identification. Unlike flow inversion where parameters were estimated along with the hydraulic head, dispersivities were not estimated. Dispersivities were assumed known as prior information for inversion using values considered representative at the inversion grid scale. Parameter estimation in transport inversion requires measured solute mass fluxes – both advective and dispersive components – which are not generally available from the field. Advancement in measurement technology could extend the capability of the current technique to joint parameter and state estimation, similar to the fluid flow inversion problems. Future work will also investigate joint flow and transport inversion by conditioning the LAS to both hydraulic and concentration measurements. In this work, for computational efficiency, solute was sampled two times at each monitoring well location, which allows the recovery of the plume history over time. This amount of data, however, is not sufficient to recover the source release history, for which denser measurements-in-time are likely needed. Future work will aim to address this for which computational efficiency is key. Moreover, uncertainty of the flow field, which gives rise to uncertainty in inferring the plume history, will also be investigated.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge the financial support by the NSF Hydrological Sciences (EAR-1702078). We thank He Huang and En-Jui Lee for comments of the parallel code. This is a theoretical work for which no data need be made available.

References

Alapati, S., Kabala, Z.J., 2000. Recovering the release history of a groundwater contaminant using a non-linear least-squares method. Hydrol. Process. 14, 1003–1016.
Ayvaz, M.T., 2016. A hybrid simulation-optimization approach for solving the areal groundwater pollution source identification problems. J. Hydrol. 538, 161–176.
Bagtzoglou, A.C., 2003. Marching-jury backward beam equation and quasi reversibility methods for hydrologic inversion: application to contaminant plume spatial distribution recovery. Water Resour. Res. 39, 1–14.
Bagtzoglou, A.C., Atmadja, J., 2005. Mathematical methods for hydrologic inversion: the case of pollution source identification. Water Pollut. Handbook Environ. Chem. 3, 65–96.
Batu, V., 2005. Applied Flow and Solute Transport Modeling in Aquifers: Fundamental Principles and Analytical and Numerical Methods. CRC Press.
Bear, J., Cheng, A.H.D., 2010. Modeling Groundwater Flow and Contaminant Transport. Springer Science & Business Media.
Chadalavada, S., Datta, B., Naidu, R., 2012. Optimal identification of groundwater pollution sources using feedback monitoring information: a case study. Environ. Forensics 13, 140–153.
Gelhar, L.W., Welty, C., Rehfeldt, K.R., 1992. A critical review of data on field-scale dispersion in aquifers. Water Resour. Res. 28 (7), 1955–1974.
Huang, H., Dennis, J.M., Wang, L., Chen, P., 2013. A scalable parallel LSQR algorithm for solving large-scale linear system for tomographic problems: a case study in seismic tomography. Procedia Comput. Sci. 18, 581–590.
Irsa, J., Zhang, Y., 2012. A direct method of parameter estimation for steady state flow in heterogeneous aquifers with unknown boundary conditions. Water Resour. Res. 48 (9). https://doi.org/10.1029/2011WR011756.
Jiao, J.Y., Zhang, Y., 2014a. Two-dimensional inversion of confined and unconfined aquifers under unknown boundary condition. Adv. Water Resour. 65, 43–57.
Jiao, J.Y., Zhang, Y., 2014b. A method based on local approximate solutions (LAS) for inverting transient flow in heterogeneous aquifer. J. Hydrol. 514, 145–149.
Jiao, J.Y., Zhang, Y., 2015a. Tensor hydraulic conductivity estimation for heterogeneous aquifers under unknown boundary conditions. Groundwater 53, 293–304. https://doi.org/10.1111/gwat.12202.
Jiao, J.Y., Zhang, Y., 2015b. Functional parameterization for hydraulic conductivity inversion with uncertainty quantification. Hydrogeol. J. 23 (3), 597–610.
Li, S., Zhang, Y., Ma, Y.Z., 2018. Comment on "A comparative study of reservoir modeling techniques and their impact on predicted performance of fluvial-dominated deltaic reservoirs". AAPG Bull. 102 (8), 1659–1663.
Michalak, A.M., Kitanidis, P.K., 2004. Estimation of historical groundwater contaminant distribution using the adjoint state method applied to geostatistical inverse modeling. Water Resour. Res. 40, 1–14.
Mirghani, B.Y., Zechman, E.M., Ranjithan, R.S., Mahinthakumar Kumar, G., 2012. Enhanced simulation-optimization approach using surrogate modeling for solving inverse problems. Environ. Forensics 13, 348–363.
National Research Council, 2013. Alternatives for Managing the Nation's Complex Contaminated Groundwater Sites. National Academies Press.
Neupauer, R.M., Borchers, B., Wilson, J.L., 2000. Comparison of inverse methods for reconstructing the release history of a groundwater contamination source. Water Resour. Res. 36, 2469.
Perkins, T.K., Johnston, O.C., 1963. A review of diffusion and dispersion in porous media. SPE J. 3 (01), 70–84.
Prakash, O., Datta, B., 2013. Sequential optimal monitoring network design and iterative spatial estimation of pollutant concentration for identification of unknown groundwater pollution source locations. Environ. Monit. Assess. 185, 5611–5626.
Seni, S.J., 1980. Sand-Body Geometry and Depositional Systems, Ogallala Formation. The University of Texas at Austin, Texas, pp. 36 Bureau of Economic Geology Report of Investigations No. 105.
Skaggs, T.H., Kabala, Z.J., 1994. Recovering the release history of a groundwater contaminant. Water Resour. Res. 30, 71–79.
Sun, A.Y., 2007. A robust geostatistical approach to contaminant source identification. Water Resour. Res. 43, 1–12.
Sun, A.Y., Painter, S.L., Wittmeyer, G.W., 2006. A robust approach for iterative contaminant source location and release history recovery. J. Contam. Hydrol. 88, 181–196.
Wang, Dongdong, 2014. Physically Based Stochastic Inversion and Parameter Uncertainty Assessment on a Confined Aquifer with A Highly Scalable Parallel Solver (M.S. Thesis). Department of Geology and Geophysics, University of Wyoming, pp. 136.
Woodbury, A.D., Ulrych, T.J., 1996. Minimum relative entropy inversion: theory and application to recovering the release history of a groundwater contaminant. Water Resour. Res. 32, 2671–2681.
Yeh, H.D., Lin, C.C., Chen, C.F., 2016. Reconstructing the release history of a groundwater contaminant based on AT123D. J. Hydroenviron. Res. 13, 89–102.
Zhang, Y., Irsa, J., Jiao, J., 2014. Three-dimensional aquifer inversion under unknown boundary conditions. J. Hydrol. 509, 416–429.
Zheng, C., 2010. MT3DMS v5.3 Supplemental User's Guide, Technical Report to the U.S. Army Engineer Research and Development Center. Department of Geological Sciences, University of Alabama, pp. 51.