Advances in Water Resources 109 (2017) 14–28

Contents lists available at ScienceDirect

Advances in Water Resources

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

Maximizing the value of pressure data in saline aquifer characterization

Seonkyoo Yoon[a,b], John R. Williams[a,b], Ruben Juanes[a,c], Peter K. Kang[a,d,*]

[a] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[b] MIT GeoNumerics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[c] Department of Earth, Atmospheric and Planetary Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA
[d] Center for Water Resources Cycle, Korea Institute of Science and Technology (KIST), Seoul, South Korea

ARTICLE INFO

Article history:
Received 25 April 2017
Revised 31 July 2017
Accepted 31 August 2017
Available online 1 September 2017

Keywords:
Managed aquifer recharge
Density-dependent flow
Inverse modeling
Ensemble Kalman filter
Value of information
Permeability estimation

ABSTRACT

The injection and storage of freshwater in saline aquifers for the purpose of managed aquifer recharge is an important technology that can help ensure sustainable water resources. As a result of the density difference between the injected freshwater and ambient saline groundwater, the pressure field is coupled to the spatial salinity distribution, and therefore experiences transient changes. The effect of variable density can be quantified by the mixed convection ratio, which is a ratio between the strength of two convection processes: free convection due to the density differences and forced convection due to hydraulic gradients. We combine a density-dependent flow and transport simulator with an ensemble Kalman filter (EnKF) to analyze the effects of freshwater injection rates on the value-of-information of transient pressure data for saline aquifer characterization. The EnKF is applied to sequentially estimate heterogeneous aquifer permeability fields using real-time pressure data. The performance of the permeability estimation is analyzed in terms of the accuracy and the uncertainty of the estimated permeability fields as well as the predictability of breakthrough curve arrival times in a realistic push-pull setting. This study demonstrates that injecting fluids at a rate that balances the two characteristic convections can maximize the value of pressure data for saline aquifer characterization.

1. Introduction

Accurate estimates of hydrogeological parameters in subsurface flow and solute transport models are critical for making predictions and managing aquifer systems. The process of estimating model input parameters, such as permeability and porosity, from observational data is often referred to as an inverse problem. Over the past few decades, various inversion methods have been proposed for groundwater modeling, and current methods are advanced enough to handle stochastic, nonlinear, and large-dimensional problems (Carrera et al., 2005; Fienen et al., 2008; Hochstetler et al., 2016; McLaughlin and Townley, 1996; Oliver and Chen, 2011; Yeh, 1986; Zhou et al., 2014). The ensemble Kalman filter (EnKF) is one such method that has gained popularity for aquifer characterization because it is easy to implement and can efficiently incorporate real-time data from a monitoring system, allowing for dynamic data assimilation (Aanonsen et al., 2009; Zhou et al., 2014). The first application of the EnKF to subsurface modeling problems was in petroleum engineering (Geir et al., 2005; Gu and Oliver, 2005; Nævdal et al., 2002); it has since been successfully extended to groundwater applications (Chen and Zhang, 2006).

The first groundwater application of the EnKF was in using groundwater flow information, such as hydraulic head data, to estimate permeability fields (Chen and Zhang, 2006; Hendricks Franssen and Kinzelbach, 2008; Tong et al., 2010). However, in a constant density groundwater flow, pressure data alone are often not sufficient to accurately estimate permeability fields; accurate estimation requires time-dependent pumping tests (Cardiff et al., 2013; 2012; Li et al., 2005) or additional data sets, such as tracer transport data (Kang et al., 2016b; Lee and Kitanidis, 2014; Li et al., 2012a; Zhang et al., 2014). The EnKF has been successfully used to incorporate multiple data sets for permeability characterization in constant density groundwater flow (Li et al., 2012a; Liu et al., 2008; Schöniger et al., 2012; Xu and Gómez-Hernández, 2016; Xu et al., 2013; Zhou et al., 2011). However, there are few inverse modeling studies of heterogeneous permeability fields in a scenario with variable-density groundwater flow and solute transport; this scenario is important for coastal aquifers experiencing

* Corresponding author.
E-mail address: pkkang@kist.re.kr (P.K. Kang).

http://dx.doi.org/10.1016/j.advwatres.2017.08.019

seawater intrusion and for managed aquifer recharge (MAR) applications in saline aquifers (Bastani et al., 2010; Kang et al., 2017b; Pool et al., 2015).

As seawater intrusion and freshwater shortages intensify, MAR is becoming an attractive technology for many coastal saline aquifers worldwide (Simmons, 2005). The coupling between fluid pressure and the spatial salinity distribution is significant in variable-density flow because the spatial salinity distribution determines the spatial fluid density distribution (Massmann et al., 2006; Simmons, 2005; Simmons et al., 2001; Ward et al., 2007; Werner et al., 2013; Zuurbier et al., 2014). This coupling between the salinity-controlled, density-driven flow and the salinity evolution leads to a time-dependent pressure; consequently, transient pressure data can be more informative for estimating aquifer permeability than in density-invariant cases (Carrera et al., 2010). Although many studies have shown the density effects on groundwater flow (Beinhorn et al., 2005; LeBlanc et al., 1991; Müller et al., 2010; Shakas et al., 2017; Vereecken et al., 2000), the variable-density effect on the value of pressure data has not been systematically studied. The first attempt to exploit this property for saline aquifer characterization was made by Kang et al. (2017b), who estimated the heterogeneous permeability field of a saline aquifer using fluid pressure data from an observational network consisting of multiple wells with pressure gauges at multiple depths. For a fixed freshwater injection rate, the authors showed that the quality of the inverse estimation does indeed improve as the density contrast between injected freshwater and the initial saline groundwater increases.

Ward et al. (2007) showed that the significance of variable-density effects during injection depends on the mixed convection ratio, which is a ratio between two characteristic types of convection: free convection due to density contrast, and forced convection due to a hydraulic gradient. For a given saline aquifer, typically there is little control over free convection because the site-specific ambient groundwater salinity determines the density contrast between injected freshwater and the ambient groundwater. However, forced convection can be controlled by human operations such as injection; thus the mixed convection ratio can be engineered by changing the freshwater injection rate.

The goal of this study is to systematically investigate how the freshwater injection rate impacts the usefulness of transient pressure data for saline aquifer characterization. To simulate a saline aquifer system where flow occurs due to the density difference between the ambient saline groundwater and injected freshwater, we developed a 2D density-dependent flow and transport model. An EnKF with covariance localization and inflation was then employed to sequentially estimate heterogeneous aquifer permeability fields using real-time pressure data. The performance of the permeability estimation was analyzed in terms of the accuracy and the uncertainty of the estimated permeability fields, and in terms of the ability of the model to predict breakthrough curve arrival times in a push-pull flow configuration not used during the estimation. The main contribution of this study is in elucidating the density effects on the value-of-information in pressure data over wide range of mixed convection regimes. To the best of knowledge, this is also the first study applying EnKF to a saline coastal aquifer system. Although this analysis was conducted for a coastal saline aquifer domain, the results are widely applicable to aquifer management and other subsurface applications in which density-driven flow is important, such as CO_2 storage and sequestration, seawater intrusion, and MAR in brackish/saline aquifers.

In Section 2 we describe the theoretical framework of mixed convection analysis for variable-density aquifer problems. In Section 3 we present the numerical model for simulating variable-density flow and transport, followed by a description of the ensemble-based data assimilation algorithm of the EnKF with covariance localization and inflation. In Section 4 we present three synthetic case studies with different types of permeability fields and monitoring networks under various mixed convection regimes. Finally, we summarize our conclusions and guidelines for future work in Section 5.

2. Mixed convection analysis

We examine a standard aquifer domain known as Henry's problem (Henry, 1964), which has been used to develop analytical and numerical approaches for considering variable-density effects (Abarca et al., 2007; Abd-Elhamid and Javadi, 2011; Frind, 1982; Huyakorn et al., 1987; Lee and Cheng, 1974; Pool and Carrera, 2011; Rastogi et al., 2004; Segol et al., 1975). Fig. 1 shows a schematic illustration of the aquifer domain and boundary conditions.

The aquifer is initially fully saturated with saline groundwater. Freshwater is injected from the domain's left boundary, while a hydrostatic pressure distribution is imposed on the right boundary. In the aquifer, fluid flow is initiated by the hydraulic gradient caused by the freshwater injection; this flow is the forced convection. For forced convection, a characteristic velocity can be defined as

$$v_{\text{forced}} = \frac{Q}{B\phi}, \tag{1}$$

where ϕ is the porosity, B is the aquifer depth in the z direction, and Q is the freshwater injection rate into a cross section of height B and unit thickness. The density difference between the injected freshwater and the ambient saline groundwater also contributes to the fluid flow; this flow is the free convection. For free convection, a characteristic velocity can also be defined as

$$v_{\text{free}} = \frac{k\Delta\rho g}{\mu\phi}, \tag{2}$$

where k is the mean permeability, $\Delta\rho$ is the density difference between the injected freshwater and initial groundwater, g is the gravitational constant, and μ is the dynamic viscosity of the fluid.

Ward et al. (2007) found that the importance of density effects depends on the interplay between forced and free convection. They introduced the mixed convection ratio, M, a dimensionless number defined as the ratio of the characteristic velocity of free convection due to density contrast to the characteristic velocity of forced convection due to freshwater injection:

$$M = \frac{v_{\text{free}}}{v_{\text{forced}}} = \frac{k\Delta\rho g B}{\mu Q}. \tag{3}$$

Mixed convective regimes can be characterized according to the mixed convection ratio. When $M \sim 1$, free and forced convection are balanced and the two characteristic velocities are approximately equal. Forced convection dominates the flow in the system when $M \ll 1$, and free convection dominates when $M \gg 1$. The tilt of the freshwater-saltwater interface increases with increases in the mixed convection ratio. Note that when there is no density difference between the injected and ambient fluids, $M = 0$.

For a given saline aquifer, we do not have control over free convection which is determined by the site-specific ambient groundwater salinity. Therefore, the freshwater injection rate determines how important the effects of density variations are, as represented by the mixed convection ratio. In order to systematically investigate how the freshwater injection rate impacts the usefulness of transient pressure data for heterogeneous permeability estimation, we use pressure data to estimate heterogeneous permeability fields in different mixed convective regimes. In the next section, we describe a forward numerical model for simulating variable-density flow and transport, and we develop a data assimilation model based on the EnKF to sequentially estimate heterogeneous permeability fields.

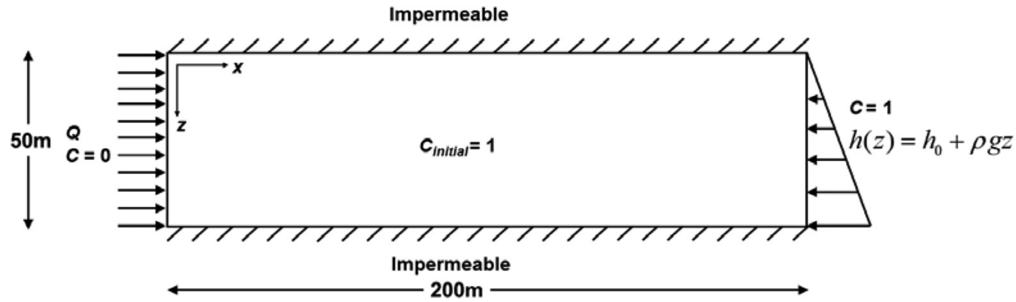

Fig. 1. Schematic illustration of the confined saline aquifer domain with boundary and initial conditions; this represents a typical coastal aquifer system. Freshwater is injected at the left boundary, where we impose a constant-flux boundary condition; a seawater hydrostatic boundary condition is imposed at the right boundary.

Table 1
Model input parameters.

Parameter	Symbol	Value	Unit
Aquifer length	L	200	[m]
Aquifer thickness	B	50	[m]
Grid size	Δx, Δy, Δz	1, 1, 1	[m]
Mean permeability	$E[\ln k]$	-23	[ln (m^2)]
Variance or ln k	$\sigma^2_{\ln k}$	0.25, 0.5, 1, 2, 3	[ln ^{2}(m^2)]
Spatial correlation model (variogram)		Gaussian[a], spherical[b]	
Effective porosity	ϕ	0.3	[-]
Longitudinal dispersivity	β_L	1	[m]
Transverse dispersivity	β_T	$\beta_L/10 = 0.1$	[m]
Fluid dynamic viscosity	μ	10^{-3}	[kg/m/s]
Freshwater density	ρ_0	1000	[kg/m^3]
Molecular diffusivity	D_0	10^{-9}	[m^2/s]
Fresh water concentration	c_0	0	[kg/kg]
Ambient saline water concentration	c_1	0.035	[kg/kg]
Number of ensembles	N_e	300	[-]
Covariance inflation factor	ω	1.01	[-]

[a] Gaussian variogram: $\gamma = 1 - \exp\left(\frac{-3h^2}{a^2}\right)$, where h and a denote lag distance and range, respectively.

[b] Spherical variogram: $\gamma = \begin{cases} \left[1.5\left(\frac{h}{a}\right) - 0.5\frac{h}{a}^3\right], & \text{for } h \le a \\ 1, & \text{otherwise} \end{cases}$. The range a is 30m directed 20° from the x-axis and 15 m directed 110° from the x-axis.

3. Methods

3.1. Numerical model of variable-density groundwater flow and transport

Fig. 1 shows a schematic for the synthetic field, and model input parameters are given in Table 1. We inject freshwater into a saline aquifer from the left boundary, simulating a fully screened well. The domain size is 200 m $\times$ 50 m, and we assign no-flow boundary conditions at the top and bottom boundaries to simulate a confined aquifer. When the ratio of the Rayleigh number (Ra), which compares buoyancy and dispersive forces, to the density difference ratio, $\alpha = \frac{\rho_{\max} - \rho_0}{\rho_0}$, is much greater than one ($\frac{Ra}{\alpha} \gg 1$), the variable density groundwater flow and transport can be described using the Boussinesq approximation (Landman and Schotting, 2007); this approximation is valid for realistic scenarios in confined saline MAR sites. The governing equations for variable-density groundwater flow and transport with the Boussinesq approximation are (Elenius et al., 2012; Hidalgo and Carrera, 2009; Hidalgo et al., 2012; Kang et al., 2017b; Landman and Schotting, 2007; Riaz et al., 2006; Szulczewski and Juanes, 2013):

$$\nabla \cdot \mathbf{u} = 0 \tag{4a}$$

$$\mathbf{u} = -\frac{\mathbf{k}}{\mu}(\nabla p - \rho(c) g \mathbf{z}) \tag{4b}$$

$$\phi \frac{\partial c}{\partial t} + \nabla \cdot (\mathbf{u} c - \phi \mathbf{D}_{\text{eff}} \nabla c) = 0. \tag{4c}$$

These governing equations consist of the continuity equation, Darcy's law, and the advection–dispersion equation, where $\mathbf{k}$ is the permeability field, ρ is the fluid density, and $\mathbf{D}_{\text{eff}}$ is the effective dispersion tensor. The Scheidegger–Bear dispersion model is used to obtain the dispersion tensor: $\phi D^{ij}_{\text{eff}} = (\phi D_0 + \beta_T |\mathbf{u}|)\delta_{ij} + (\beta_L - \beta_T)\frac{u_i u_j}{|\mathbf{u}|}$, where D_0 is the molecular diffusivity, $|\mathbf{u}|$ is the magnitude of the Darcy velocity, β_{L} is the longitudinal dispersivity, and β_{T} is the transverse dispersivity. Density is a linear function of concentration, $\rho = \rho_0 + \frac{\partial \rho}{\partial c}(c - c_0)$ where $\frac{\partial \rho}{\partial c} = 700$ [kg/m^3] and ρ_0 is the density of freshwater (Voss and Souza, 1987). c is the concentration of solute as a mass fraction of dissolved salt in water (mass of dissolved salt per unit mass of fluid), and $c_0 = 0$ [kg/kg] for injected freshwater and $c_1 = 0.035$ [kg/kg] for ambient saline groundwater. The aquifer is initially fully saturated with saline groundwater, and we start to inject freshwater at $t = 0$.

The boundary conditions are given by:

$$\mathbf{u} \cdot \mathbf{n}(x = 0, z, t) = v_{\text{forced}} \tag{5a}$$

$$\mathbf{u} \cdot \mathbf{n}(x, z = 0 \text{ or } B, t) = 0 \tag{5b}$$

$$p(x = L, z, t) = \rho_{\text{sea}} g z \tag{5c}$$

where $\mathbf{n}$ is the outward unit normal to the boundary. We inject freshwater at a constant flow rate from the left boundary and assign a seawater hydrostatic pressure boundary condition at the right boundary (Voss and Souza, 1987). For all simulations, the model domain is discretized into a grid of 200 $\times$ 50 cells; each cell is a square element with dimensions $\Delta x = \Delta z = 1$ m. We solve for the pressure field using a finite volume method with a two-point flux approximation (TPFA), then solve for the concentration field using a finite volume method with an upwind scheme; we

Table 2
Mixed convection ratio M and corresponding injection rates Q.

M [-]	Q [m^2/d]	Duration [day]
0	38125.7	0.06
0.001	38125.7	0.06
0.01	3812.6	0.6
0.1	381.3	6
1	38.1	60
5	7.6	300
10	3.8	600

then integrate in time using an explicit forward Euler scheme (LeVeque, 2002). The detailed model input parameters are given in Table 1.

We varied the mixed convection ratio, $M = 0 - 10$, by varying the freshwater injection rate (Table 2) to investigate its impact on the inversion results. $M = 0$ corresponds to a case with no density contrast between the injected water and the groundwater, which results in steady-state groundwater flow. To confirm the generality of our inversion results, we study log-normal permeability fields with different levels of heterogeneity and two different variogram models (Gaussian and spherical).

3.2. Inverse estimation with ensemble Kalman filter

We apply an ensemble Kalman filter (EnKF) to estimate the heterogeneous permeability fields by assimilating real-time pressure data from a spatially sparse monitoring system. The EnKF is based on the Kalman filter, which is an optimal solution to a recursive Bayesian update problem for a linear, stochastic state–space system with additive Gaussian errors (Kalman, 1960). When a system is linear, an exact propagation of the state covariance matrix in time is possible, so the optimal solution provides analytical formulas for updating the mean and covariance of the system state vector. However, most Bayesian update problems, including subsurface flow and transport modeling, are nonlinear; thus the covariance matrix cannot be analytically updated. The ensemble Kalman filter (EnKF) is a Monte-Carlo implementation of the Kalman filter. The EnKF circumvents the nonlinearity problem by replacing the state covariance with a sample covariance; this is commonly called an ensemble covariance and is computed from ensemble realizations of the state vector (Evensen, 1994; 2003; 2009).

The EnKF has been successfully applied to groundwater problems. Nowak (2009) provided a theoretical basis, derived from the principles of unbiasedness and minimum error variance, for using the EnKF to estimate geostatistical model parameters that are conditional on transient model state variables. Li et al. (2012a) applied the EnKF assimilating concentration and piezometric head data to estimate not only model parameters such as hydraulic conductivity and porosity but also state variables such as pressure and concentration. This allows for assessing the predictability of flow and transport behavior during the analysis. The application of the EnKF has also been successfully extended to more general cases such as non-Gaussian systems (Li et al., 2012b; Schöniger et al., 2012; Xu and Gómez-Hernández, 2016; Xu et al., 2013; Zhou et al., 2011).

The EnKF algorithm starts from an initial ensemble of aquifer models generated from *a priori* geostatistical assumptions. Each aquifer model is represented by a state vector Y, comprising the model parameters and the state variables. Because pressure data are used to update the modeled permeability, the state vector in our study is composed of the permeability, pressure, and salinity at each grid cell, $Y = [\ln \mathbf{k}^T, \mathbf{p}^T, \mathbf{c}^T]^T$. The ensemble of all state vectors is collected in a matrix as

$$\mathbf{Y} = \left[Y_1, Y_2, \ldots, Y_{N_e}\right], \tag{6}$$

where N_e denotes the total number of ensemble members. The EnKF begins the assimilation procedure when the first set of observation data becomes available. The EnKF then updates the ensemble matrix $\mathbf{Y}$ to match the measurements at every assimilation step by using the Kalman formula to correct the ensemble matrix. This update for each ensemble member is given by

$$Y_j^u = Y_j^f + \mathbf{C}_Y^f \mathbf{H}^T \left(\mathbf{H}\mathbf{C}_Y^f \mathbf{H}^T + \mathbf{C_d}\right)^{-1} \left(\mathbf{d}_{\text{obs}} + \epsilon_j - \mathbf{H}Y_j^f\right). \tag{7}$$

Here $j = [0, \ldots, N_e]$ is the index of the individual ensemble member, the superscript u indicates the updated value, and the superscript f indicates a forecasted value. The observation available at the current assimilation step is $\mathbf{d}_{\text{obs}}$, $\mathbf{C_d}$ is the covariance matrix of the measurement noise, and ϵ_j is an observation error with zero mean and covariance $\mathbf{C_d}$. $\mathbf{H}$ is a matrix operator that selects predicted variables from the state vector, $\mathbf{C}_Y\mathbf{H}^T$is the cross-covariance between all the state variables and the predicted observations, and $\mathbf{H}\mathbf{C}_Y\mathbf{H}^T$ is the auto-covariance of the predicted observations. The covariance of the state vector is approximated using the standard statistical formula

$$\mathbf{C}_Y^f = \frac{1}{N_e - 1} \sum \left(Y_j^f - \bar{Y}^f\right)\left(Y_j^f - \bar{Y}^f\right)^T, \tag{8}$$

where $\bar{Y}^f$ denotes the mean of the state vectors. The update step in Eq. (7) can yield unphysical parameter values, such as salt concentrations outside the range 0–1. When that happens, the unphysical concentration values are reset to the closest bound, 0 or 1. The state vectors updated by Eq. (7) are simulated forward in time to the next data assimilation point; this is the forecast step. Then, the predicted ensemble matrix is again updated using the data through Eq. (7). This recursive procedure continues until all the measurements have been assimilated. Fig. 2 shows a flow chart of this sequential data assimilation via the EnKF.

For large-scale inverse problems, the covariance in Eq. (8) can be repeatedly underestimated over a sequence of updates if the ensemble size is small (Furrer and Bengtsson, 2007; Hendricks Franssen and Kinzelbach, 2008). This problem can be alleviated by covariance inflation which is implemented by multiplying the deviation of the state vector from the ensemble mean by an amount, ω, larger than one (Anderson, 2007; Hamill et al., 2001):

$$Y_j^{f,\text{infl}} = \omega \times \left(Y_j^f - \overline{Y}^f\right) + Y_j^f. \tag{9}$$

In this study, we applied one percent inflation ($\omega = 1.01$), as in Hamill et al. (2001).

Small ensemble sizes also often result in spurious correlations between state components that are physically far apart. To avoid this problem, localization schemes have been proposed (Chang et al., 2010; Houtekamer and Mitchell, 2001; Sun et al., 2009; Tong et al., 2012; Xu et al., 2013). The key idea is to taper the covariance matrix according to the distance between grid points. Many functions to reduce the sample covariance between spatially distant components have been proposed (Bergemann and Reich, 2010; Campbell et al., 2010; Chen and Oliver, 2010; Constantinescu et al., 2007; Devegowda et al., 2010; Gaspari and Cohn, 1999; Greybush et al., 2011; Houtekamer and Mitchell, 2001; Nan and Wu, 2011). We use the fifth-order function of Gaspari and Cohn (1999), which replaces the ensemble Kalman filter update in Eq. (7) with the following localized ensemble Kalman filter (LEnKF) update:

$$Y_j^u = Y_j^f + \left[\left(\tau(d) \circ \mathbf{C}_Y^f\right)\mathbf{H}^T\right]\left[\mathbf{H}\left(\tau(d) \circ \mathbf{C}_Y^f\right)\mathbf{H}^T + \mathbf{C_d}\right]^{-1} \times \left(\mathbf{d}_{\text{obs}} + \epsilon_j - \mathbf{H}Y_j^f\right), \tag{10}$$

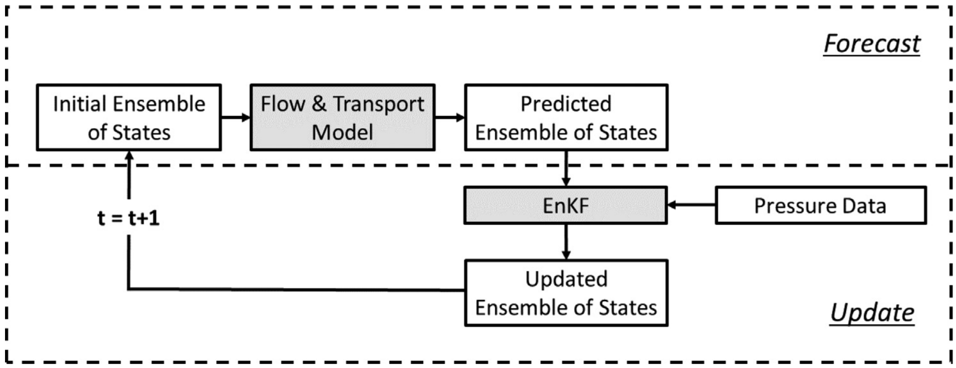

Fig. 2. Flow chart showing data assimilation via the ensemble Kalman filter.

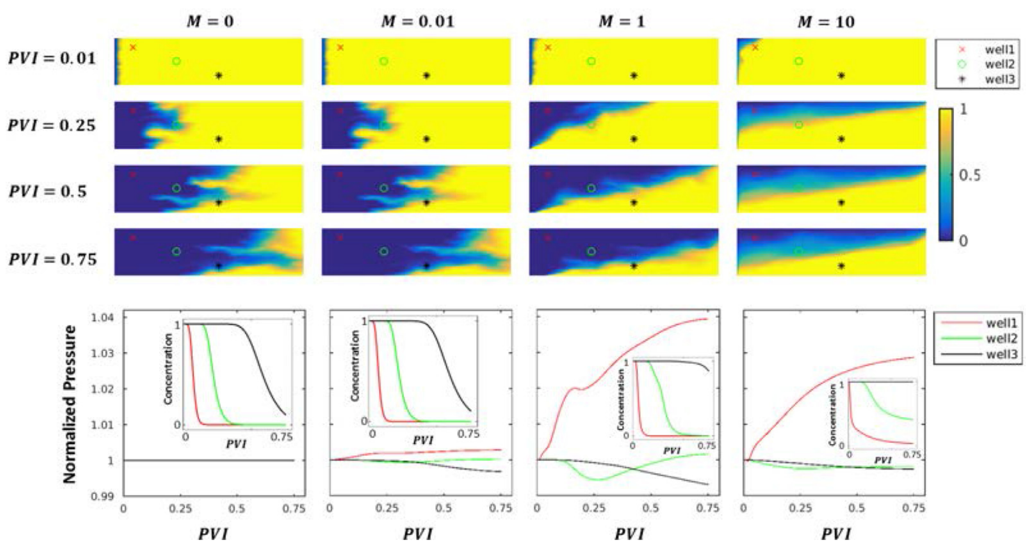

Fig. 3. Flow and transport simulation results for different values of the mixed convection ratio *M*. The 16 upper panels show the evolution of the injected freshwater plume for different pore volumes injected of freshwater (PVI); the colormap indicates the salinity of the water. The lower four panels show the normalized pressure measurements as a function of PVI (which is a proxy for time) at three different observation locations; these locations are indicated on the upper panels. The pressure values are normalized by the initial pressure value, and the variation of the normalized pressure value is largest at $M = 1$. The bottom panels also include, as insets, the breakthrough curves (concentration as a function of PVI) at each of the three wells.

where $\circ$ is the Schur product. The tapering function $\tau(d)$ is defined as

$$\tau(d) = \begin{cases} -\frac{1}{4}\left(\frac{d}{r}\right)^5 + \frac{1}{2}\left(\frac{d}{r}\right)^4 + \frac{5}{8}\left(\frac{d}{r}\right)^3 - \frac{5}{3}\left(\frac{d}{r}\right)^2 + 1, & 0 \le d \le r, \\ \frac{1}{12}\left(\frac{d}{r}\right)^5 - \frac{1}{2}\left(\frac{d}{r}\right)^4 + \frac{5}{8}\left(\frac{d}{r}\right)^3 + \frac{5}{3}\left(\frac{d}{r}\right)^2 & \\ \quad -5\left(\frac{d}{r}\right) + 4 - \frac{2}{3}\left(\frac{d}{r}\right)^{-1}, & r \le d \le 2r, \\ 0, & 2r \le d, \end{cases} \tag{11}$$

where d is the distance between two points and $2r$ is the range beyond which the tapering function yields zero correlation. In this study, we apply the LEnKF with covariance inflation to estimate heterogeneous permeability fields.

4. Results and discussion

4.1. Impact of variable-density flow on pressure data

We first consider the effect of variable density on the dynamics of the injected freshwater. Fig. 3 shows the results of forward-model simulations in a heterogeneous permeability field. Concentration maps at different values of the pore volume injected (PVI) are shown for different values of the mixed convection ratio, *M*. The mixed convection ratio is varied by changing the freshwater injection rate (Table 2). As Fig. 3 shows, the mixed convection ratio has a significant impact on both plume spreading and the time evolution of the fluid pressure data. For cases dominated by free convection ($M > 1$), the interface between the injected freshwater and the ambient saline water tilts significantly; freshwater cannot sweep the entire domain because free convection dominates the fluid flow. At $M = 1$, density effects are still evident, but the injection rate is high enough that the freshwater sweeps the whole domain. In cases where forced convection dominates, the effect of density differences diminishes, and injected freshwater effectively sweeps the entire domain.

We find that the transient fluid pressure behavior is very sensitive to variable-density effects; this implies that the freshwater injection rate might control how informative pressure measurements are. When there is no density contrast between the injected and defending fluids, the pressure field is insensitive to the salt concentration field; it stays constant in time. For $M > 0$, we clearly observe changes in pressure over time due to the density-driven flow. Note that the relative pressure change for $M = 1$ is significantly larger than for $M = 0.01$. This is because the free convection component that causes the pressure change becomes dominant as the mixed convection ratio increases. However, the relative pressure change becomes smaller at $M = 10$ compared to the $M = 1$ case because

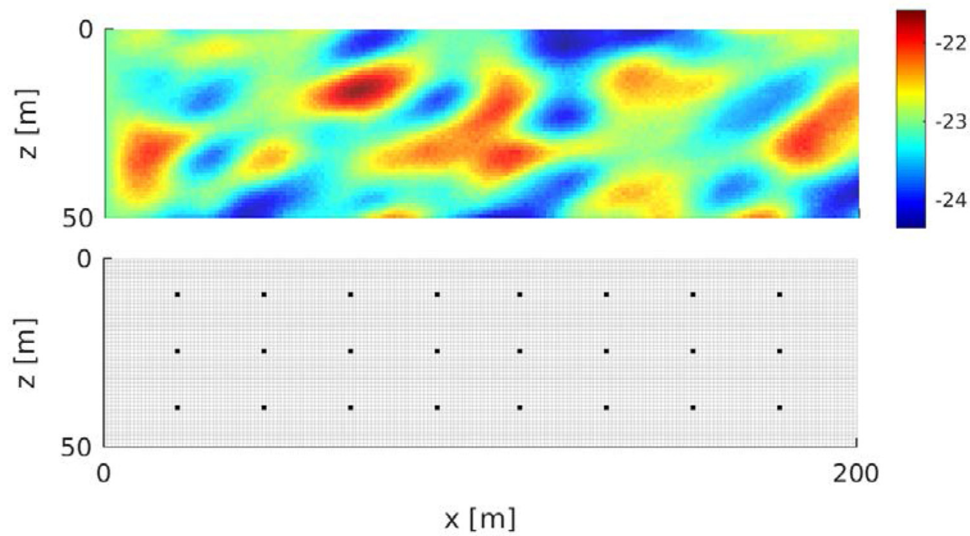

Fig. 4. The reference (true) log-permeability field (top panel) and locations of pressure measurements in the low heterogeneity case study (bottom panel). The log-permeability field is assumed to be multi-Gaussian with a log-permeability mean $E\left[\ln k\right] = -23$ and a log-permeability variance $\sigma^2_{\ln k} = 0.25$. The field is defined on a 200×50 grid with cells of size $1\,\text{m} \times 1\,\text{m}$. There are 24 measurement points, whose log-permeability values are assumed to be known and equal to the mean value $E\left[\ln k\right] = -23$. The spatial correlation structure is modeled by a Gaussian variogram..

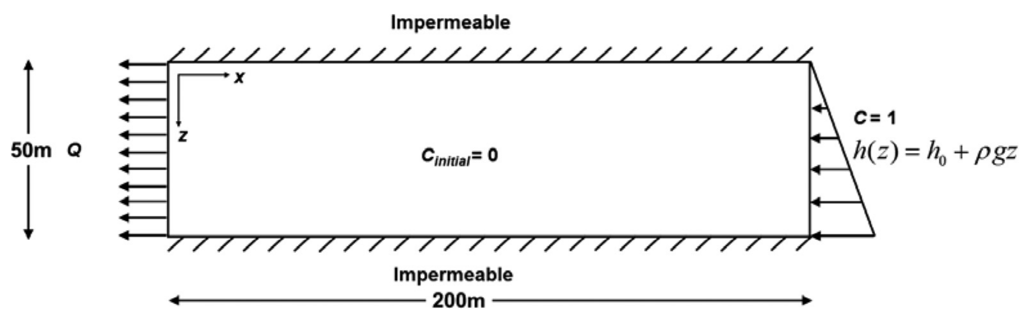

Fig. 5. Conceptual model for transport predictability tests showing initial and boundary conditions. The case study is constructed to simulate a seawater intrusion scenario. We impose a seawater hydrostatic pressure boundary condition at the right boundary, and freshwater is produced at a constant rate at the left boundary.

of the poor sweeping efficiency at high M. Qualitative observations of the forward simulation results indicate that the pressure field is most sensitive at $M \approx 1$.

The coupling between the pressure field and the spatial salinity distribution causes nontrivial pressure changes; these changes occur *before* the injected fluid reaches the pressure sampling point. This shows that changes in the freshwater–saline groundwater interface influence the pressure distribution globally as well as locally. The degree of pressure change is larger when the freshwater sweeps through the observation point. Having established that the transient pressure field is affected by variable-density flow and the freshwater injection rate, the key question is whether this pressure field can provide information about subsurface permeability structures and, if so, how the value of the pressure data can be optimized.

4.2. Inversion results

To investigate how variable-density flow affects the use of transient pressure data, we compare the permeability fields estimated via EnKF for different freshwater injection rates to the known reference fields. Fig. 4 shows an example of the reference log-permeability field along with the locations of pressure observations points in a sparse monitoring system. There are eight observation wells with multilevel groundwater monitoring system that give pressure data at three discrete levels (Einarson and Cherry, 2002; Pickens et al., 1978). We use measured pressure values at these 24 data sampling points during a freshwater injection experiment to estimate the heterogeneous permeability field.

The quality of the permeability estimation is rigorously analyzed using four different metrics. The first inverse estimation error measurement is based on Euclidean distance (l_2-norm),

$$e = \left\| \ln \mathbf{k}^{\text{true}} - \ln \overline{\mathbf{k}} \right\|, \tag{12}$$

where $\mathbf{k}^{\text{true}}$ is a vector of the true permeability values at every grid cell and $\overline{\mathbf{k}}$ is a vector of the mean values of the updated ensemble of permeability fields.

Second, the inverse estimation is assessed by mapping accuracy, which is the fraction of correctly estimated grid cells with regard to the true permeability field (Yoon and McKenna, 2012). To define the criteria for the correct estimation, we first calculate the difference between maximum and minimum values of the true log-permeability as $\delta_{\ln k^{\text{true}}} = \max(\ln \mathbf{k}^{\text{true}}) - \min(\ln \mathbf{k}^{\text{true}})$. Then, a log-permeability estimate of a grid cell is counted as correct when the difference between the true and estimated log-permeability values is less than a certain threshold, which (for example) can be 10 percent of the difference $\delta_{\ln k^{\text{true}}}$, as $0.1\delta_{\ln k^{\text{true}}}$. Note that the mapping accuracy evaluates the fraction of accurately estimated grid cells within an estimated permeability field, while the aforementioned measure based on Euclidean distance assesses the overall similarity between permeability fields.

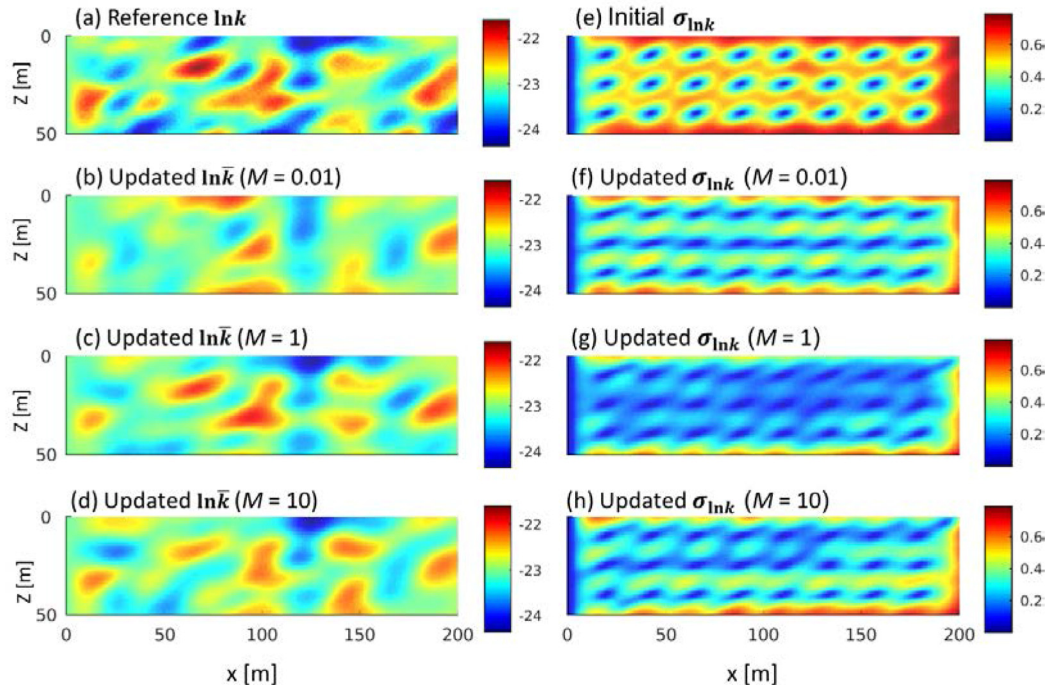

Fig. 6. Panel (a): the true permeability field in the low heterogeneity case study. Panels (b-d): mean of the estimated permeability fields after the final update for the cases with $\beta = 1$ and $M = 0.01$ (b), $M = 1$ (c), and $M = 10$ (d). Panel (e): standard deviation of the initial ensemble of permeability fields. Panels (f–h): standard deviation of the estimated permeability fields after the final update for the cases with $\beta = 1$ and $M = 0.01$ (f), $M = 1$ (g), and $M = 10$ (h).

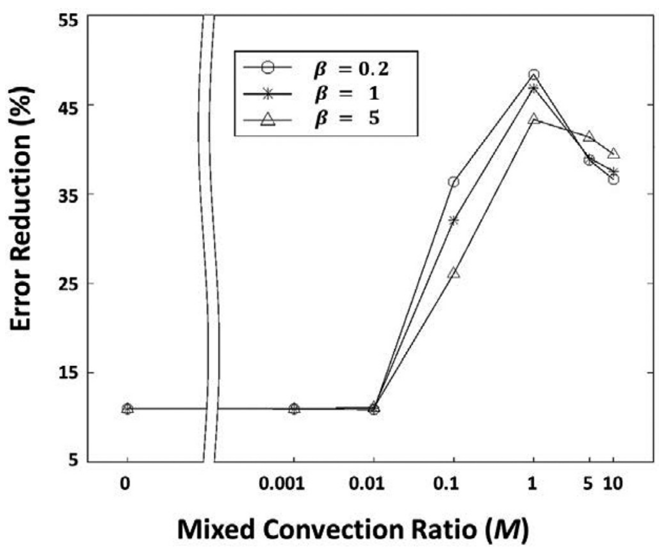

Fig. 7. The reduced error as a function of the mixed convection ratio M for three different values of β in the low heterogeneity case study. The reduced error is defined as $\frac{e^{\text{initial}}-e}{e^{\text{initial}}}$, where e^{initial} is the error of the initial permeability ensemble. This measure represents the accuracy of the estimated permeability field, which is maximized at the balanced mixed convection ratio of one.

Third, we analyze the uncertainty of the inverse estimation. The posterior covariance of the log-permeability field obtained after the pressure data $\mathbf{d}_{\text{obs},\,t}$ at time t is assimilated via Eq. (10) can be approximated as $\text{Cov}(\ln \mathbf{k}|\mathbf{d}_{\text{obs},t}) = \frac{1}{N_e-1}\sum(\ln \mathbf{k}^u_j - \overline{\ln \mathbf{k}^u})(\ln \mathbf{k}^u_j - \overline{\ln \mathbf{k}^u})^T$, where $\overline{\ln \mathbf{k}^u}$ is the mean of the updated ensemble of log-permeability fields. The uncertainty of the inverse estimation can then be quantified as

$$\Lambda_t = \text{tr}[\text{Cov}(\ln \mathbf{k}|\mathbf{d}_{\text{obs},t})], \tag{13}$$

where the trace provides a scalar measure of the posterior variance of the updated log-permeability (Neuman et al., 2012). Note that the value of additional measurements $\mathbf{d}_{\text{obs},\,t}$ in reducing the uncertainty can be quantified as $\Lambda_t - \Lambda_{t-1}$, where Λ_{t-1} is the trace of the updated covariance matrix in the previous time window (Dai et al., 2016).

Finally, we analyze what in practice is the most important quality of an estimated permeability field: its ability to faithfully predict flow and transport under different flow scenarios. The estimated permeability fields are used to simulate a seawater intrusion scenario; the aquifer domain is initially filled with freshwater and then invaded with seawater from the right boundary (Fig. 5). The salinity arrival times at the pumping well on the left boundary are measured and compared with the true arrival times obtained from the known permeability field.

We now present three synthetic case studies with different types of permeability fields (different levels of heterogeneity and two different variogram models) and monitoring networks.

4.2.1. Case 1: low heterogeneity

The first case study assumes a log-permeability field following a Gaussian variogram with mean $E[\ln k] = -23$; this corresponds to approximately $k = 10^5$ millidarcy. The variance of the log-permeability is $(\sigma^2_{\ln k} = 0.25)$, which is similar to those of low-heterogeneity natural geological formations such as Borden, Ontario $(\sigma^2_{\ln k} = 0.29)$ and Cape Cod, Massachusetts$(\sigma^2_{\ln k} = 0.24)$ (Garabedian et al., 1991; Hess et al., 1992; Mackay et al., 1986). Note that we assume that the $\ln \mathbf{k}$ is multi-Gaussian with preknown spatial statistics such as the mean and covariance. There are 24 measurement points, whose log-permeability values are assumed to be known and fixed to the mean value of $E[\ln k] = -23$. The assumption of the known values of the log-permeability at the measurement points is applied by enforcing the initial ensemble state vectors to have the known value at the corresponding well points. The enforcement is carried out by the conditional sequential Gaussian simulation (Remy et al., 2009). Such a

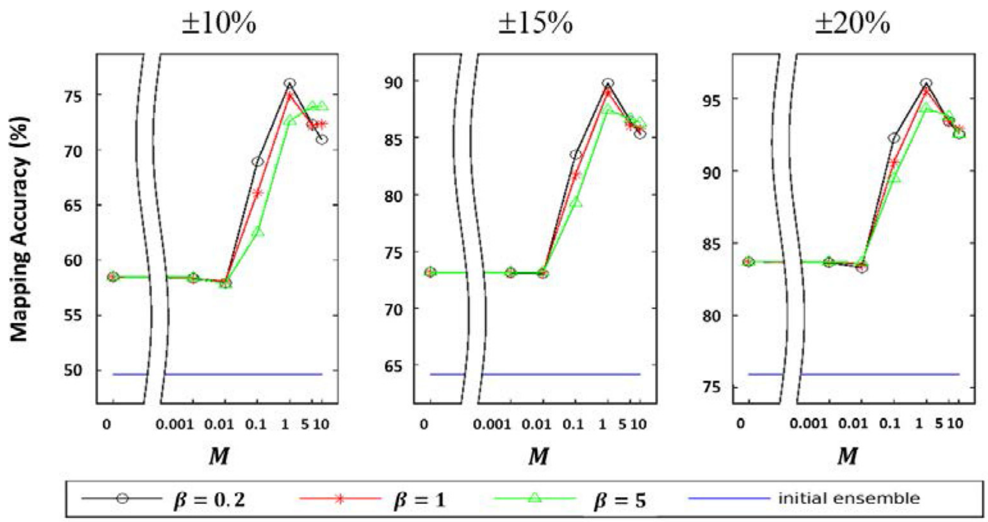

Fig. 8. The mapping accuracy of the finally updated ensemble mean of log-permeability fields in the low heterogeneity case study as a function of the mixed convection ratio M for three different criteria (10%, 15%, 20%). The mapping accuracy is the fraction of correctly estimated grid cells with regard to the true permeability field. A log-permeability estimate of a grid cell is counted correct when the difference between the true and estimated log-permeability values is less than a certain threshold.

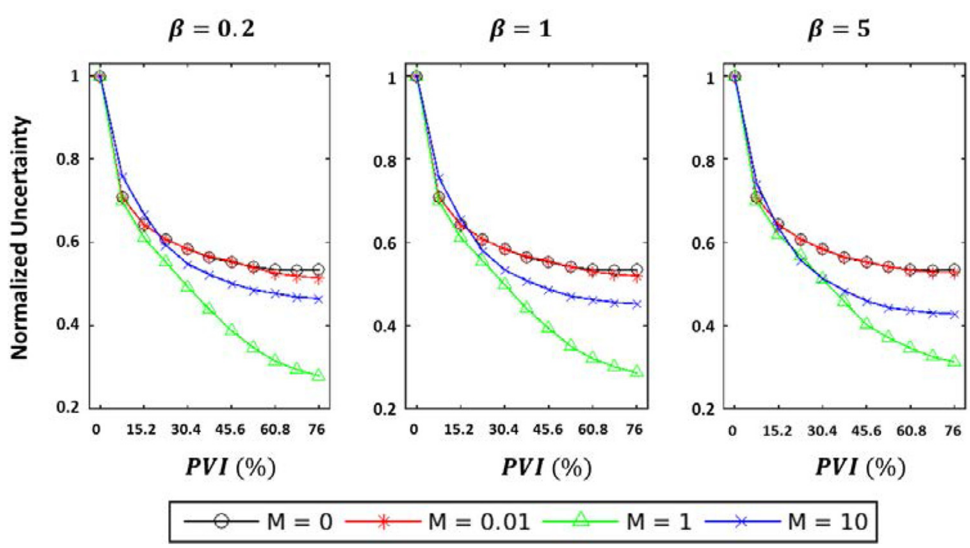

Fig. 9. The time evolution of the normalized uncertainty measure, $\frac{\Lambda_t}{\Lambda_0}$, for $M = [0, 0.01, 1, 10]$ and $\beta = [0.2, 1, 5]$ as a function of pore volume injected (PVI) for the low heterogeneity case study. Ten data assimilation steps are conducted for each case, and the largest uncertainty reduction is for the balanced mixed convection ratio of one.

construction of the permeability fields with conditional data at measurement points is used assuming laboratory or in-situ permeability measurements are available (Kitanidis and Vomvoris, 1983; Xu and Gómez-Hernández, 2015; Zhou et al., 2011). The spatial correlation structure is modeled by a Gaussian variogram with a range of 30 m in the direction 20° from the x-axis and 15 m in the direction 110° from x-axis. The same statistics are used to generate 300 initial permeability ensemble members. Note that the initial ensemble does not contain the true permeability field.

For various values of the mixed convection ratio, $M = [0, 0.001, 0.01, 0.1, 1, 5, 10]$, and dispersivity, $\beta = [0.2, 1, 5]$, the LEnKF is applied to assimilate the measured pressure data and estimate the permeability field. The total amount of injected water for all cases is fixed to be 2287 m^3, which represents 76 % of the total pore volume of the domain. The injection rate for each value of M is summarized in Table 2. There are ten updates for each scenario, and each update occurs when 10 % of the total freshwater has been injected since the previous update. At the time of each update, the pressure measurements from the monitoring points shown in Fig. 4 are assimilated.

The mean and standard deviation of the estimated permeability fields after the last update for $M = [0.01, 1, 10]$ and $\beta = 1$ are shown in Fig. 6. Qualitatively, the $M = 1$ case shows the most accurately estimated permeability field with the lowest uncertainty. Kang et al. (2017b) reported that inversion estimates performed significantly better when variable-density effects are present compared to constant density cases. This is because the pressure field is steady in the case of constant density, thereby providing limited information for permeability estimation. However, they did not report the decreased estimation accuracy at $M = 10$ compared to $M = 1$ that is apparent in Fig. 6.

We rigorously assess the performance of permeability estimation using the four measures discussed in Section 4.2. We normalize the error estimates, e, in Eq. (12): $\frac{e^{\text{initial}} - e}{e^{\text{initial}}}$, where e^{initial} is the error of the initial permeability ensemble. The normalized error estimates can be interpreted to be the reduced error, as shown in Fig. 7. The performance of the inversion improves as the mixed

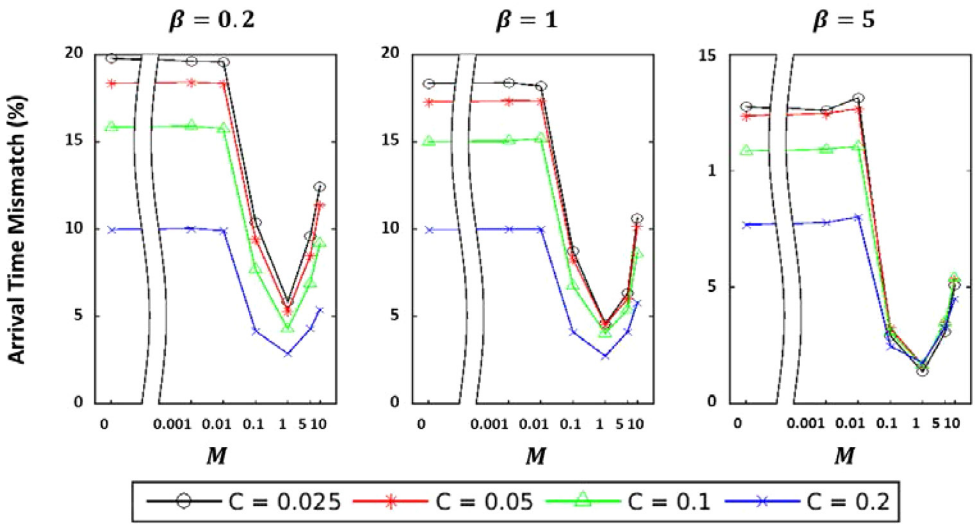

Fig. 10. The results of the transport predictability tests for three different values of β in the low heterogeneity case study. The arrival time mismatch, $\frac{|t_{\mathrm{true}}-t_{\mathrm{predict}}|}{t_{\mathrm{true}}}$, is minimized at a balanced mixed convection ratio of one. The results are consistent for arrival times of four different concentration values: 2.5%, 5%, 10%, and 20% of the seawater concentration.

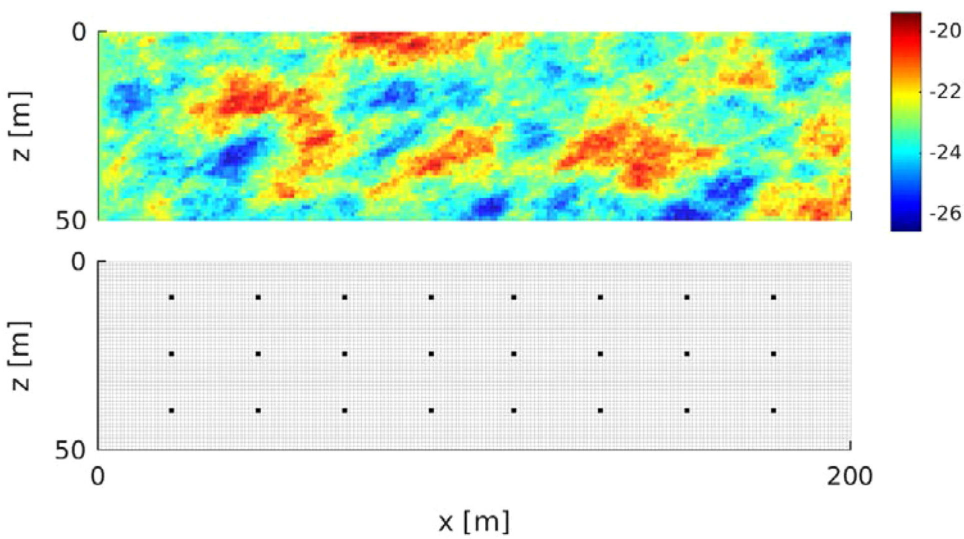

Fig. 11. The reference (true) log-permeability field and locations of pressure measurements in the high heterogeneity case study. The log-permeability field is assumed to be multi-Gaussian with a log-permeability mean $E[\ln k] = -23$ and a log-permeability variance $\sigma^2_{\ln k} = 1$. The field is defined on a 200×50 grid with cells of size $1\,\mathrm{m} \times 1\,\mathrm{m}$. There are 24 measurement points whose log-permeability values are assumed to be known and equal to the mean value $E[\ln k] = -23$. The spatial correlation structure is modeled using a spherical variogram.

convection ratio increases because the variable-density effects increase the value of the pressure data. When density variations are important, the pressure field is coupled with the spatial salinity distribution via Eq. (4b); this implies that the transient pressure measurements also contain salt transport information (Kang et al., 2017b). However, the error reduction begins to decrease when the mixed convection ratio increases above one; this corresponds to cases where the free convection is larger than the forced convection, and the interface between the injected freshwater and the initial saline water tilts significantly due to the density contrast between the fluids (see Fig. 3). The tilting prevents the injected freshwater from sweeping the whole aquifer domain, and density-driven flow occurs only in the upper areas that the freshwater plume passes through. Consequently, pressure change originating from the displacement of saltwater by freshwater occurs only in a restricted area. This diagnosis is also confirmed by the estimated mapping accuracy shown in Fig. 8. The mapping accuracy increases as the variable-density effects increase, and decreases after reaching the maximum at the balanced mixed convection regime ($M = 1$).

The values of the uncertainty measure, Λ_t, computed from Eq. (12) at every assimilation time point for each value of M and β, are normalized by the uncertainty of the initial ensemble, Λ_0. Fig. 9 shows the evolution of the normalized uncertainty measures as a function of PVI. For all combinations of M and β, the first update gives the largest reduction in uncertainty. For density-invariant cases ($M = 0$), the uncertainty decreases and the curves plateau earlier than the cases where significant variable-density effects exist. The largest and continued uncertainty reduction is observed for the balanced mixed convection regime ($M = 1$), implying that transient pressure data continuously provides useful information for uncertainty reduction. When we further increase the mixed convection ratio, the trend reverses and the uncertainty reduction is not as large.

We also evaluated the practical performance of the inversion in terms of transport predictions. For this purpose, we designed

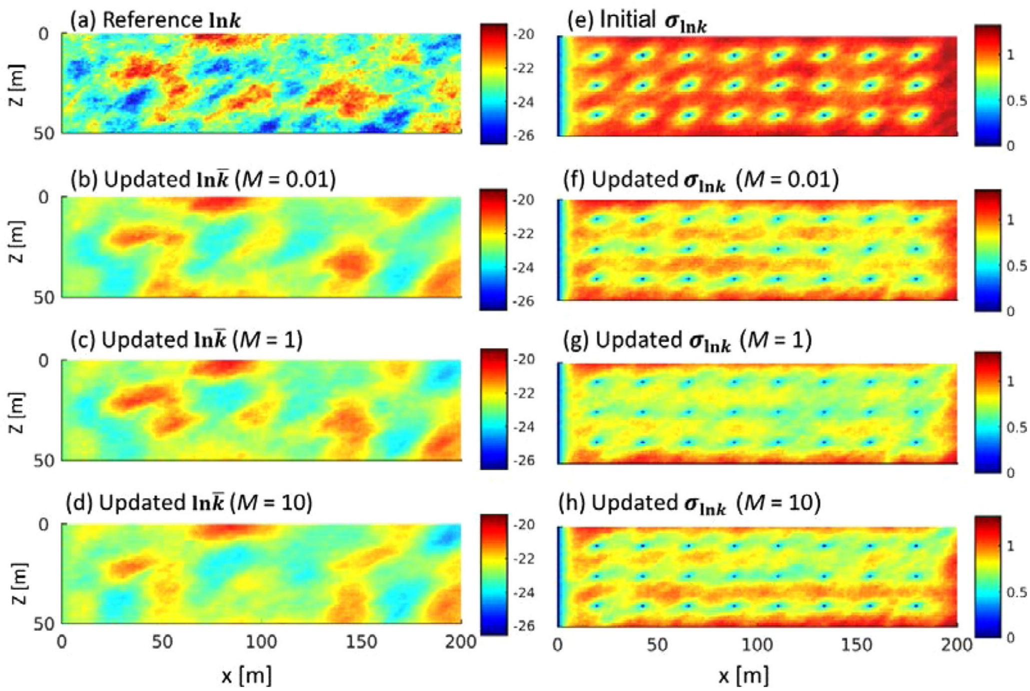

Fig. 12. Panel (a): the true permeability field in the high heterogeneity case ($\sigma^2_{\ln k} = 1$) study. Panels (b-d): mean of the estimated permeability fields after the final update for the cases with $\beta = 1$ and $M = 0.01$ (b), $M = 1$ (c), and $M = 10$ (d). Panel (e): standard deviation of the initial ensemble of permeability fields. Panels (f-h): standard deviation of the estimated permeability fields after the final update for the cases with $\beta = 1$ and $M = 0.01$ (f), $M = 1$ (g), and $M = 10$ (h).

a saltwater intrusion experiment; the aquifer domain is initially fully saturated with freshwater and a saltwater intrusion is simulated by producing freshwater from the well on the left boundary at a constant rate of $38\,\mathrm{m}^3$/day (Fig. 5). The pumping allows for 76% of the total pore volume of the domain to be withdrawn in 60 days, and the breakthrough curves of salinity at the pumping well are measured. We measure breakthrough curves for all cases ($M = [0, 0.001, 0.01, 0.1, 1, 5, 10]$ and $\beta = [0.2, 1, 5]$). The difference between the true arrival time t_{true} and the predicted time t_{predict} is normalized by the true arrival time as $\frac{|t_{\text{true}} - t_{\text{predict}}|}{t_{\text{true}}}$, and the results are shown in Fig. 10. The predictability of arrival times is maximized at $M = 1$ for all cases, which is consistent with the accuracy and uncertainty reduction analysis shown above.

4.2.2. Case 2: high heterogeneity

We conduct the same analysis for permeability fields with higher degrees of heterogeneity ($\sigma^2_{\ln k}$ up to 3) to determine whether our identification of an optimum mixed convection regime for permeability estimation can be generalized. As a representative case, we first present the $\sigma^2_{\ln k} = 1$ case. The spatial correlation structure is modeled using a spherical variogram, which generates fields that are less smooth than with a Gaussian variogram model. Fig. 11 shows the true log-permeability field along with the monitoring system.

The mean and standard deviation of the estimated permeability fields for $M = [0.01, 1, 10]$ and $\beta = 1$ are shown in Fig. 12. Qualitatively, the $M = 1$ case again has the most accurately estimated permeability field with the least uncertainty. We confirmed with the four measures that the value of the pressure data is maximized at the balanced mixed convection regime also for $\sigma^2_{\ln k} = 1$ with the spherical variogram case. We present, for brevity, only the error reduction and the time evolution of the normalized uncertainty measure. The estimated error reductions for different values of the mixed convection ratio and dispersivity are shown in Fig. 13. Similar to the low-heterogeneity case, the error reduction of the inverse estimation initially improves as the mixed convection ratio

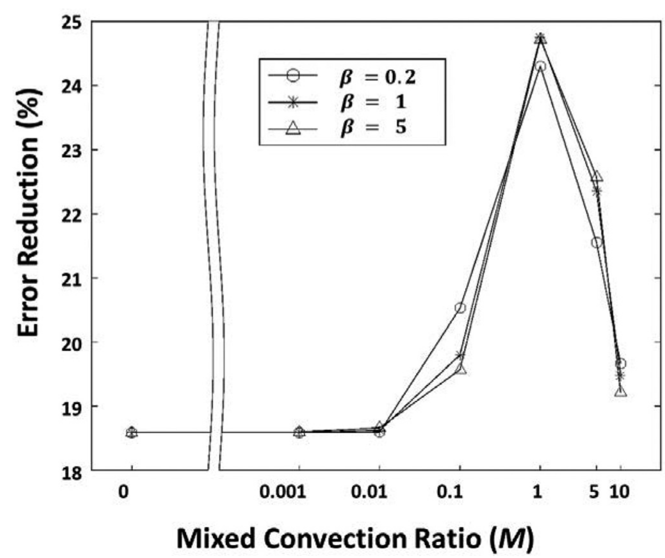

Fig. 13. The reduced error as a function of the mixed convection ratio M for three different values of β in the high heterogeneity case ($\sigma^2_{\ln k} = 1$) study. The reduced error is defined as $\frac{e^{\text{initial}} - e}{e^{\text{initial}}}$, where e^{initial} is the error of the initial permeability ensemble. This represents the accuracy of the estimated permeability field, which is maximized at a balanced mixed convection ratio of one.

increases to $M = 1$; then it decreases as the mixed convection ratio becomes larger than one. Fig. 14 shows the evolution of the normalized uncertainty measure introduced in Eq. (13). For all values of M, the first update gives the biggest reduction in uncertainty. As seen in the low heterogeneity case, the largest and most continued uncertainty reduction occurs for the balanced mixed convection regime ($M = 1$).

We also performed the inverse modeling for higher values of heterogeneity ($\sigma^2_{\ln k}$ up to 3) and found that the inverse estimation is always optimum around $M \approx 1$. In Fig. 15, we show the

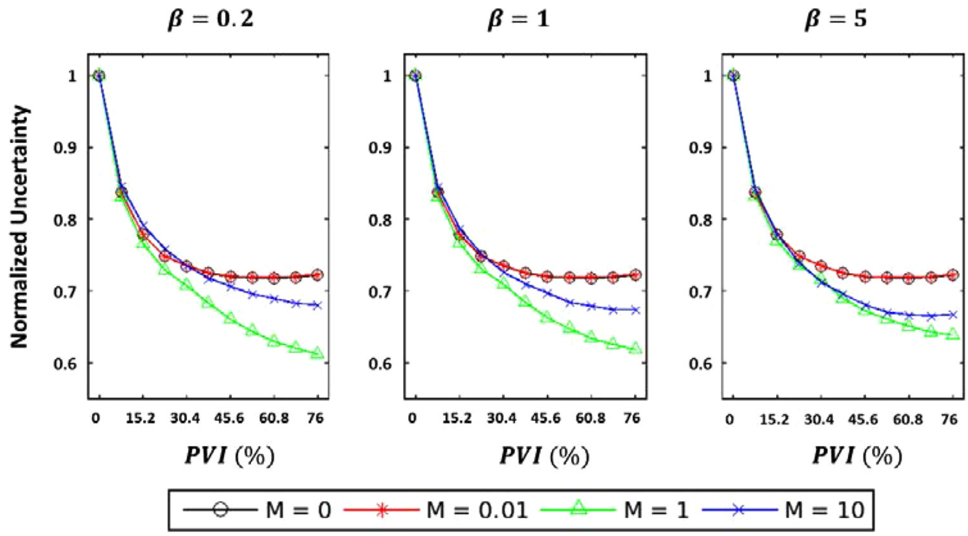

Fig. 14. The time evolution of the normalized uncertainty measure, $\frac{\Lambda_t}{\Lambda_0}$, for $M = [0, 0.01, 1, 10]$ and $\beta = [0.2, 1, 5]$ as a function of pore volume injected (PVI) for the high heterogeneity case ($\sigma^2_{\ln k} = 1$) study. Ten data assimilation steps are conducted for each case, and the largest uncertainty reduction is for the balanced mixed convection ratio of one.

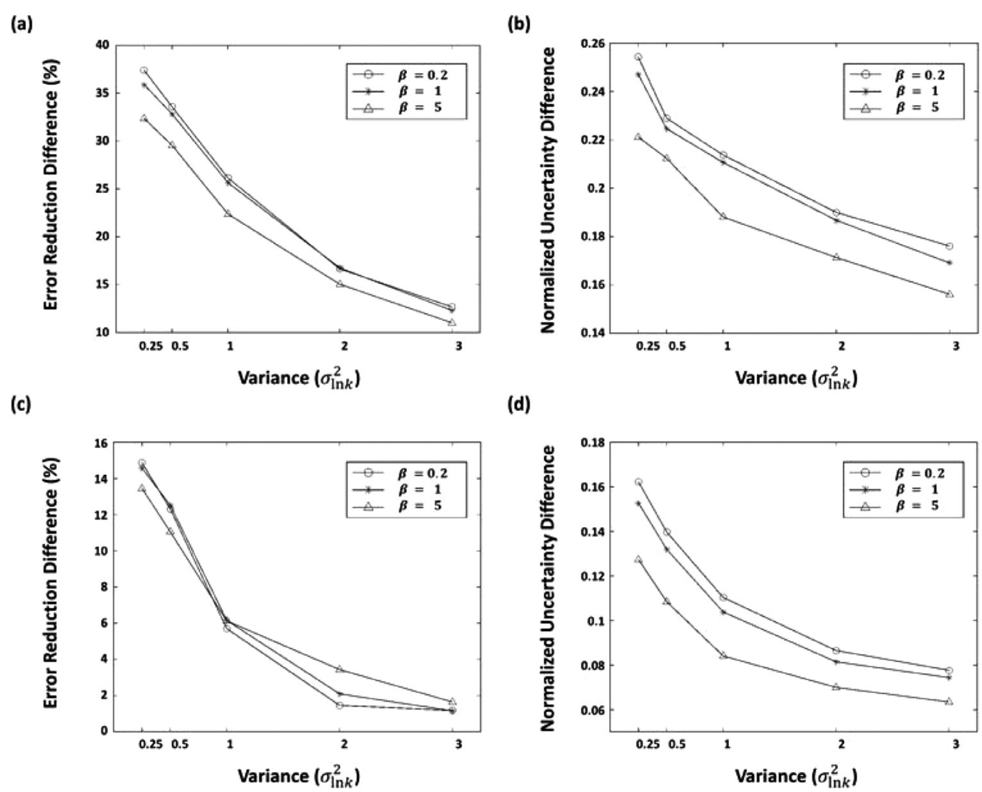

Fig. 15. Differences, or improvements, of the error reduction and the estimation uncertainty between the optimum mixed convection regime ($M = 1$) and the density-invariant case ($M = 0$) as a function of $\sigma^2_{\ln k}$. (a-b) the improvements when the Gaussian variogram is used. (c-d) the improvements when the spherical variogram is used. The estimation improvement decreases as heterogeneity increases.

differences (or improvements) of the error reduction and the estimation uncertainty between the optimum mixed convection regime ($M = 1$) and the density-invariant case ($M = 0$) for different levels of heterogeneity. We clearly observe that the estimation improvement decreases as heterogeneity increases for both Gaussian and spherical variogram cases. This might be explained by the fact that the heterogeneity causes preferential flow that dominates flow behavior (Fiori and Jankovic, 2012; Kang et al., 2016a; 2017a; Kung, 1990). This implies that the variable-density effect on flow behavior will decrease as heterogeneity increases.

4.2.3. Case 3: other settings

We provide an additional case study to demonstrate that our findings are also valid for different settings of the observation network. We test a moderately heterogeneous case with $\sigma^2_{\ln k} = 0.5$ and $E[\ln k] = -23$, as shown in Fig. 16. There are 16 measurement

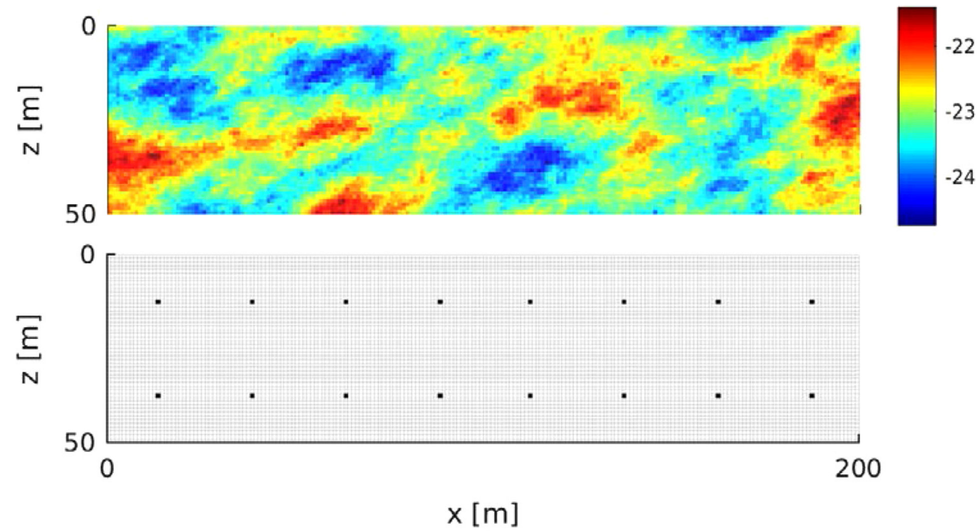

Fig. 16. The reference (true) log-permeability field and locations of pressure measurements. The log-permeability field is assumed to be multi-Gaussian with a log-permeability mean $E[\ln k] = -23$ and a log-permeability variance $\sigma^2_{\ln k} = 0.5$. The field is defined on a 200×50 grid with cells of size $1\,\mathrm{m} \times 1\,\mathrm{m}$. There are 16 measurement points (a third less than the previous cases). The log-permeability values at these points are assumed *unknown*. The spatial correlation structure is modeled using a spherical variogram.

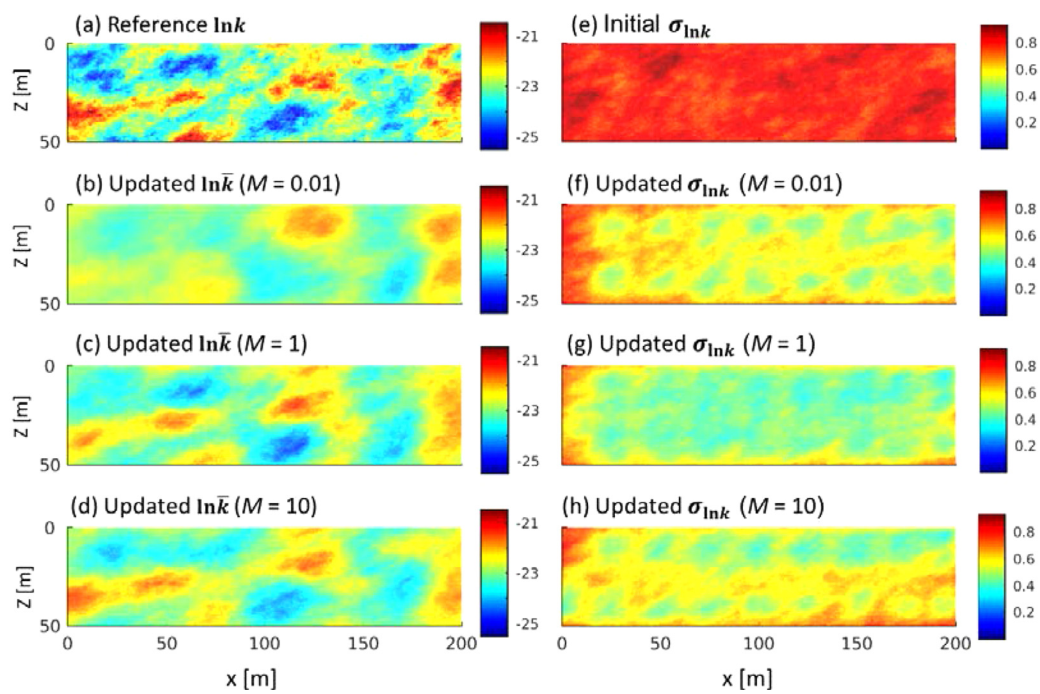

Fig. 17. Panel (a): the true permeability field in the moderate heterogeneity case study. Panels (b-d): mean of the estimated permeability fields after the final update for the cases with $\beta = 1$ and $M = 0.01$ (b), $M = 1$ (c), and $M = 10$ (d). Panel (e): standard deviation of the initial ensemble of permeability fields. Panels (f-h): standard deviation of the estimated permeability fields after the final update for the cases with $\beta = 1$ and $M = 0.01$ (f), $M = 1$ (g), and $M = 10$ (h).

points, a third less than in the previous case studies. Unlike the previous case studies, the log-permeability values at the measurement points are assumed *unknown*. The spatial correlation structure is modeled using a spherical variogram.

Fig. 17 shows the mean and standard deviation of the estimated permeability fields. Qualitatively, the $M = 1$ case again has the most accurately estimated permeability field with the least uncertainty. The optimality of the balanced mixed convection regime for maximizing the value of the pressure data is also confirmed by the four performance measures, and we present the error reduction in Fig. 18. This confirms that the optimality of the balanced mixed convection regime is valid for the case with fewer monitoring points.

We also performed a sensitivity analysis to assess the effects of the covariance inflation and localization. Three scenarios of the covariance treatments are considered: 1. No covariance treatments; 2. With inflation ($\omega = 1.01$) but no localization; 3. With localization ($r = 50$m) but no inflation. The estimated error reductions for the three different scenarios are shown in Fig. 18. The covariance localization improves the estimation accuracy, whereas the improvements by the covariance inflation are insignificant. This implies that the ensemble size of 300 is large enough to avoid the ensemble collapse, while circumventing spurious correlations between distant points by the covariance localization improves the estimation accuracy. The results confirm that the estimation accuracy is

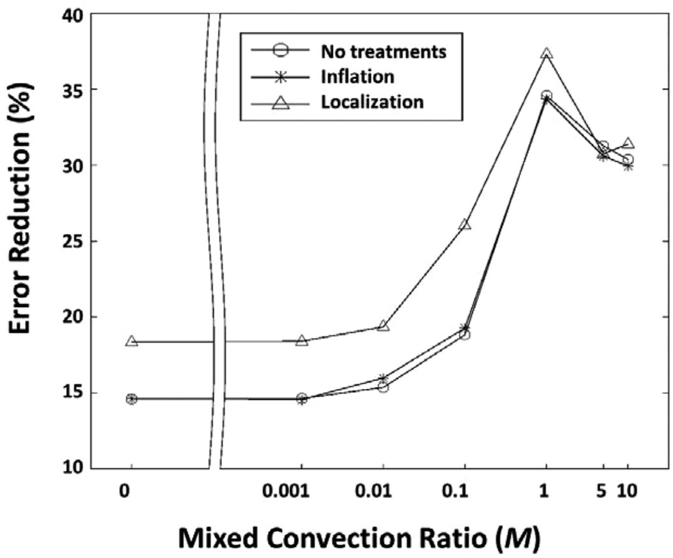

Fig. 18. The reduced error as a function of the mixed convection ratio M for three different scenarios of covariance treatments: 1. No covariance treatments; 2. With covariance inflation ($\omega = 1.01$); 3. With covariance localization ($r = 50$m). The reduced error represents the accuracy of the estimated permeability field. The result shows that the accuracy is maximized at a balanced mixed convection ratio regardless of the covariance treatments.

maximized at the balanced mixed convection regime ($M = 1$) regardless of the covariance treatments.

5. Conclusions

We have demonstrated that freshwater injection rates in saline aquifers significantly influence the value of pressure data for aquifer characterization. The fact that the pressure distribution is coupled with the density gradient means pressure measurements are more informative in variable density cases than in constant density. However, when density-driven free convection overrides forced convection, pressure data become less useful for aquifer characterization because the interface between injected freshwater and ambient saline water tilts significantly. This tilt prevents the injected freshwater from sweeping the entire aquifer domain, and variable-density flow occurs only in a limited area, making the pressure data less informative. An important finding is that the value of pressure data can be maximized when the two types of convection are balanced, corresponding to a mixed convection ratio of one. This finding is rigorously shown for different types of permeability fields and monitoring networks using four different measures: error reduction, mapping accuracy, estimation uncertainty, and transport predictability.

This study shows that mixed convection regimes should be considered in saline aquifer characterization. More specifically, this work suggests the possibility of improving aquifer characterization by enforcing a balanced mixed convection regime in the aquifer system via human operations such as managing the freshwater injection rate. In real field applications, the mixed convection regime can be modified by varying the density of the injection fluid (Shakas et al., 2017) or by varying the injection rate. Although our analysis was conducted for the coastal aquifer domain, the implications of this work might have wide applicability for aquifer management, CO_2 storage and sequestration, seawater intrusion, and MAR in coastal areas.

In this study, we only investigated the impact of variable-density effects on the use of pressure data. However, other types of measurements, such as concentration measurements or production rates from wells, can also be considered. An evaluation of the use of different types of data in saline aquifer characterization will be the subject of future work. This line of research will allow us to determine the ideal combination of data and effective monitoring protocols for aquifers with density-dependent flow.

Acknowledgments

Seonkyoo Yoon acknowledges support from the Massachusetts Institute of Technology (MIT) Energy Initiative (MITEI) Seed Fund Program. Peter K. Kang acknowledges a grant (code 17AWMP-B066761-05) from the AWMP Program funded by the Ministry of Land, Infrastructure and Transport of the Korean government and the support from Future Research Program (2E27030) funded by the Korea Institute of Science and Technology (KIST).

References

Aanonsen, S.I., Nævdal, G., Oliver, D.S., Reynolds, A.C., Vallès, B., et al., 2009. The ensemble Kalman filter in reservoir engineering–a review. SPE J. 14 (03), 393–412.
Abarca, E., Carrera, J., Sánchez-Vila, X., Dentz, M., 2007. Anisotropic dispersive Henry problem. Adv. Water Resour. 30 (4), 913–926.
Abd-Elhamid, H., Javadi, A., 2011. A density-dependant finite element model for analysis of saltwater intrusion in coastal aquifers. J. Hydrol. 401 (3), 259–271.
Anderson, J.L., 2007. An adaptive covariance inflation error correction algorithm for ensemble filters. Tellus A 59 (2), 210–224. http://dx.doi.org/10.1111/j.1600-0870.2006.00216.x.
Bastani, M., Kholghi, M., Rakhshandehroo, G.R., 2010. Inverse modeling of variable–density groundwater flow in a semi-arid area in iran using a genetic algorithm. Hydrogeol. J. 18 (5), 1191–1203.
Beinhorn, M., Dietrich, P., Kolditz, O., 2005. 3-D numerical evaluation of density effects on tracer tests. J. Contam. Hydrol. 81 (1), 89–105.
Bergemann, K., Reich, S., 2010. A localization technique for ensemble Kalman filters. Q. J. R. Meteorolog. Soc. 136 (648), 701–707.
Campbell, W.F., Bishop, C.H., Hodyss, D., 2010. Vertical covariance localization for satellite radiances in ensemble Kalman filters. Mon. Weather Rev. 138 (1), 282–290.
Cardiff, M., Bakhos, T., Kitanidis, P.K., Barrash, W., 2013. Aquifer heterogeneity characterization with oscillatory pumping: sensitivity analysis and imaging potential. Water Resour. Res. 49 (9), 5395–5410.
Cardiff, M., Barrash, W., Kitanidis, P.K., 2012. A field proof-of-concept of aquifer imaging using 3-d transient hydraulic tomography with modular, temporarily-emplaced equipment. Water Resour. Res. 48 (5). http://dx.doi.org/10.1029/2011WR011704. W05531.
Carrera, J., Alcolea, A., Medina, A., Hidalgo, J., Slooten, L.J., 2005. Inverse problem in hydrogeology. Hydrogeol. J. 13 (1), 206–222.
Carrera, J., Hidalgo, J.J., Slooten, L.J., Vázquez-Suñé, E., 2010. Computational and conceptual issues in the calibration of seawater intrusion models. Hydrogeol. J. 18 (1), 131–145.
Chang, H., Chen, Y., Zhang, D., 2010. Data assimilation of coupled fluid flow and geomechanics using the ensemble Kalman filter. SPE J. 15 (02), 382–394.
Chen, Y., Oliver, D.S., 2010. Cross-covariances and localization for EnKF in multiphase flow data assimilation. Comput. Geosci. 14 (4), 579–601.
Chen, Y., Zhang, D., 2006. Data assimilation for transient flow in geologic formations via ensemble Kalman filter. Adv. Water Resour. 29 (8), 1107–1122.
Constantinescu, E.M., Sandu, A., Chai, T., Carmichael, G.R., 2007. Ensemble-based chemical data assimilation. II: covariance localization. Q. J. R. Meteorolog. Soc. 133 (626), 1245–1256.
Dai, C., Xue, L., Zhang, D., Guadagnini, A., 2016. Data-worth analysis through probabilistic collocation-based ensemble kalman filter. J. Hydrol. 540, 488–503.
Devegowda, D., Arroyo-Negrete, E., Datta-Gupta, A., 2010. Flow relevant covariance localization during dynamic data assimilation using enkf. Adv Water Resour 33 (2), 129–145.
Einarson, M.D., Cherry, J.A., 2002. A new multilevel ground water monitoring system using multichannel tubing. Groundwater Monitoring Remediation 22 (4), 52–65.
Elenius, M.T., Nordbotten, J.M., Kalisch, H., 2012. Effects of a capillary transition zone on the stability of a diffusive boundary layer. IMA J. Appl. Math. 77 (6), 771–787.
Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99 (C5), 10143–10162.
Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. Ocean Dyn. 53 (4), 343–367.
Evensen, G., 2009. Data Assimilation: The Ensemble Kalman Filter. Springer.
Fienen, M.N., Clemo, T., Kitanidis, P.K., 2008. An interactive Bayesian geostatistical inverse protocol for hydraulic tomography. Water Resour. Res. 44 (12), W00B01.
Fiori, A., Jankovic, I., 2012. On preferential flow, channeling and connectivity in heterogeneous porous formations. Math. Geosci. 44 (2), 133–145. http://dx.doi.org/10.1007/s11004-011-9365-2.
Frind, E.O., 1982. Simulation of long-term transient density-dependent transport in groundwater. Adv. Water Resour. 5 (2), 73–88.
Furrer, R., Bengtsson, T., 2007. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. J. Multivar. Anal. 98 (2), 227–255. http://dx.doi.org/10.1016/j.jmva.2006.08.003.

Garabedian, S.P., LeBlanc, D.R., Gelhar, L.W., Celia, M.A., 1991. Large-scale natural gradient tracer test in sand and gravel, cape cod, massachusetts: 2. Analysis of spatial moments for a nonreactive tracer. Water Resour. Res. 27 (5), 911–924. http://dx.doi.org/10.1029/91WR00242.

Gaspari, G., Cohn, S.E., 1999. Construction of correlation functions in two and three dimensions. Q. J. R. Meteorolog. Soc. 125 (554), 723–757.

Geir, N., Johnsen, L.M., Aanonsen, S.I., Vefring, E.H., 2005. Reservoir monitoring and continuous model updating using ensemble Kalman filter. SPE J. 10 (1), 66–74.

Greybush, S.J., Kalnay, E., Miyoshi, T., Ide, K., Hunt, B.R., 2011. Balance and ensemble kalman filter localization techniques. Mon. Weather Rev. 139 (2), 511–522.

Gu, Y., Oliver, D.S., 2005. History matching of the PUNQ-S3 reservoir model using the ensemble kalman filter. SPE J. 10 (02), 217–224.

Hamill, T.M., Whitaker, J.S., Snyder, C., 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. Mon. Weather Rev. 129 (11), 2776–2790.

Hendricks Franssen, H., Kinzelbach, W., 2008. Real-time groundwater flow modeling with the ensemble kalman filter: joint estimation of states and parameters and the filter inbreeding problem. Water Resour. Res. 44 (9). http://dx.doi.org/10.1029/2007WR006505. W09408.

Henry, H. R., 1964. Effects of dispersion on salt encroachment in coastal aquifers. US Geological Survey Water Supply Paper 1063-C, C70–C84.

Hess, K.M., Wolf, S.H., Celia, M.A., 1992. Large-scale natural gradient tracer test in sand and gravel, cape cod, massachusetts: 3. hydraulic conductivity variability and calculated macrodispersivities. Water Resour. Res. 28 (8), 2011–2027. http://dx.doi.org/10.1029/92WR00668.

Hidalgo, J.J., Carrera, J., 2009. Effect of dispersion on the onset of convection during CO_2 sequestration. J. Fluid Mech. 640, 441–452.

Hidalgo, J.J., Fe, J., Cueto-Felgueroso, L., Juanes, R., 2012. Scaling of convective mixing in porous media. Phys. Rev. Lett. 109, 264503.

Hochstetler, D.L., Barrash, W., Leven, C., Cardiff, M., Chidichimo, F., Kitanidis, P.K., 2016. Hydraulic tomography: continuity and discontinuity of high-K and low-K zones. Groundwater 54 (2), 171–185. http://dx.doi.org/10.1111/gwat.12344.

Houtekamer, P.L., Mitchell, H.L., 2001. A sequential ensemble kalman filter for atmospheric data assimilation. Mon. Weather Rev. 129 (1), 123–137.

Huyakorn, P.S., Andersen, P.F., Mercer, J.W., White, H.O., 1987. Saltwater intrusion in aquifers: development and testing of a three-dimensional finite element model. Water Resour. Res. 23 (2), 293–312.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. J. Basic Eng. 82 (1), 35–45.

Kang, P.K., Brown, S., Juanes, R., 2016. Emergence of anomalous transport in stressed rough fractures. Earth Planet. Sci. Lett 454, 46–54.

Kang, P.K., Dentz, M., Le Borgne, T., Lee, S., Juanes, R., 2017. Anomalous transport in disordered fracture networks: spatial markov model for dispersion with variable injection modes. Adv. Water Resour. http://dx.doi.org/10.1016/j.advwatres.2017.03.024.

Kang, P.K., Lee, J., Fu, X., Lee, S., Kitanidis, P.K., Juanes, R., 2017. Improved characterization of heterogeneous permeability in saline aquifers from transient pressure data during freshwater injection. Water Resour. Res. 53 (5), 4444–4458. http://dx.doi.org/10.1002/2016WR020089.

Kang, P.K., Zheng, Y., Fang, X., Wojcik, R., McLaughlin, D., Brown, S., Fehler, M.C., Burns, D.R., Juanes, R., 2016. Sequential approach to joint flow–seismic inversion for improved characterization of fractured media. Water Resour. Res. 52 (2), 903–919. http://dx.doi.org/10.1002/2015WR017412.

Kitanidis, P.K., Vomvoris, E.G., 1983. A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. Water Resour. Res. 19 (3), 677–690. http://dx.doi.org/10.1029/WR019i003p00677.

Kung, K.-J., 1990. Preferential flow in a sandy vadose zone: 1. Field observation. Geoderma 46 (1), 51–58. http://dx.doi.org/10.1016/0016-7061(90)90006-U.

Landman, A.J., Schotting, R.J., 2007. Heat and brine transport in porous media: the oberbeck-Boussinesq approximation revisited. Transp. Porous Med. 70 (3), 355–373.

LeBlanc, D.R., Garabedian, S.P., Hess, K.M., Gelhar, L.W., Quadri, R.D., Stollenwerk, K.G., Wood, W.W., 1991. Large-scale natural gradient tracer test in sand and gravel, cape cod, Massachusetts: 1. Experimental design and observed tracer movement. Water Resour. Res. 27 (5), 895–910. http://dx.doi.org/10.1029/91WR00241.

Lee, C.-H., Cheng, R.T.-S., 1974. On seawater encroachment in coastal aquifers. Water Resour. Res. 10 (5), 1039–1043.

Lee, J., Kitanidis, P.K., 2014. Large-scale hydraulic tomography and joint inversion of head and tracer data using the principal component geostatistical approach (PCGA). Water Resour. Res. 50 (7), 5410–5427.

LeVeque, R.J., 2002. Finite Volume Methods for Hyperbolic Problems, 31. Cambridge University Press.

Li, L., Zhou, H., Gómez-Hernández, J.J., Franssen, H.-J.H., 2012. Jointly mapping hydraulic conductivity and porosity by assimilating concentration data via ensemble kalman filter. J. Hydrol. 428, 152–169.

Li, L., Zhou, H., Hendricks Franssen, H.J., Gómez-Hernández, J.J., 2012. Groundwater flow inverse modeling in non-multigaussian media: performance assessment of the normal-score ensemble kalman filter. Hydrol. Earth Syst. Sci. 16 (2), 573–590.

Li, W., Nowak, W., Cirpka, O.A., 2005. Geostatistical inverse modeling of transient pumping tests using temporal moments of drawdown. Water Resour. Res. 41 (8). http://dx.doi.org/10.1029/2004WR003874. W08403.

Liu, G., Chen, Y., Zhang, D., 2008. Investigation of flow and transport processes at the MADE site using ensemble kalman filter. Adv. Water Resour. 31 (7), 975–986.

Mackay, D.M., Freyberg, D.L., Roberts, P.V., Cherry, J.A., 1986. A natural gradient experiment on solute transport in a sand aquifer: 1. approach and overview of plume movement. Water Resour. Res. 22 (13), 2017–2029. http://dx.doi.org/10.1029/WR022i013p02017.

Massmann, G., Simmons, C., Love, A., Ward, J., James-Smith, J., 2006. On variable density surface water–groundwater interaction: a theoretical analysis of mixed convection in a stably-stratified fresh surface water–saline groundwater discharge zone. J. Hydrol. 329 (3), 390–402.

McLaughlin, D., Townley, L.R., 1996. A reassessment of the groundwater inverse problem. Water Resour. Res. 32 (5), 1131–1161.

Müller, K., Vanderborght, J., Englert, A., Kemna, A., Huisman, J.A., Rings, J., Vereecken, H., 2010. Imaging and characterization of solute transport during two tracer tests in a shallow aquifer using electrical resistivity tomography and multilevel groundwater samplers. Water Resour. Res. 46 (3). http://dx.doi.org/10.1029/2008WR007595. W03502.

Nævdal, G., Mannseth, T., Vefring, E.H., 2002. Near-well reservoir monitoring through ensemble Kalman filter (SPE-75235). SPE/DOE Improved Oil Recovery Symposium.

Nan, T., Wu, J., 2011. Groundwater parameter estimation using the ensemble kalman filter with localization. Hydrogeol. J. 19 (3), 547–561.

Neuman, S.P., Xue, L., Ye, M., Lu, D., 2012. Bayesian analysis of data-worth considering model and parameter uncertainties. Adv. Water Resour. 36, 75–85.

Nowak, W., 2009. Best unbiased ensemble linearization and the quasi-linear kalman ensemble generator. Water Resour. Res. 45 (4). http://dx.doi.org/10.1029/2008WR007328. W04431.

Oliver, D.S., Chen, Y., 2011. Recent progress on reservoir history matching: a review. Comput. Geosci. 15 (1), 185–221.

Pickens, J.F., Cherry, J.A., Grisak, G.E., Merritt, W.F., Risto, B.A., 1978. Multilevel device for groundwater sampling and piezometric monitoring. Ground Water 16 (5), 322–327.

Pool, M., Carrera, J., 2011. A correction factor to account for mixing in Ghyben-Herzberg and critical pumping rate approximations of seawater intrusion in coastal aquifers. Water Resour. Res, 47 (5). http://dx.doi.org/10.1029/2010WR010256. W05506.

Pool, M., Carrera, J., Alcolea, A., Bocanegra, E.M., 2015. A comparison of deterministic and stochastic approaches for regional scale inverse modeling on the mar del plata aquifer. J. Hydrol. 531, 214–229.

Rastogi, A., Choi, G.W., Ukarande, S., 2004. Diffused interface model to prevent ingress of sea water in multi-layer coastal aquifers. J. Spatial Hydrol. 4 (2).

Remy, N., Boucher, A., Wu, J., 2009. Applied Geostatistics with SGeMS: A User's Guide. Cambridge University Press.

Riaz, A., Hesse, M., Tchelepi, H.A., Orr, F.M., 2006. Onset of convection in a gravitationally unstable diffusive boundary layer in porous media. J. Fluid Mech. 548, 87–111.

Schöniger, A., Nowak, W., Hendricks Franssen, H.-J., 2012. Parameter estimation by ensemble kalman filters with transformed data: approach and application to hydraulic tomography. Water Resour. Res. 48 (4). http://dx.doi.org/10.1029/2011WR010462. W04502.

Segol, G., Pinder, G.F., Gray, W.G., 1975. A Galerkin-finite element technique for calculating the transient position of the saltwater front. Water Resour. Res. 11 (2), 343–347.

Shakas, A., Linde, N., Baron, L., Selker, J., Gerard, M.-F., Lavenant, N., Bour, O., Le Borgne, T., 2017. Neutrally buoyant tracers in hydrogeophysics: field demonstration in fractured rock. Geophys. Res. Lett. 44 (8), 3663–3671. http://dx.doi.org/10.1002/2017GL073368. 2017GL073368.

Simmons, C.T., 2005. Variable density groundwater flow: from current challenges to future possibilities. Hydrogeol. J. 13 (1), 116–119.

Simmons, C.T., Fenstemaker, T.R., Sharp, J.M., 2001. Variable-density groundwater flow and solute transport in heterogeneous porous media: approaches, resolutions and future challenges. J. Contam. Hydrol. 52 (1), 245–275.

Sun, A.Y., Morris, A.P., Mohanty, S., 2009. Sequential updating of multimodal hydrogeologic parameter fields using localization and clustering techniques. Water Resour. Res. 45 (7). http://dx.doi.org/10.1029/2008WR007443. W07424.

Szulczewski, M.L., Juanes, R., 2013. The evolution of miscible gravity currents in horizontal porous layers. J. Fluid Mech. 719, 82–96.

Tong, J., Hu, B.X., Yang, J., 2010. Using data assimilation method to calibrate a heterogeneous conductivity field conditioning on transient flow test data. Stochastic Environ. Res. Risk Assess. 24 (8), 1211–1223.

Tong, J., Hu, B.X., Yang, J., 2012. Assimilating transient groundwater flow data via a localized ensemble Kalman filter to calibrate a heterogeneous conductivity field. Stochastic Environ. Res. Risk Assess. 26 (3), 467–478.

Vereecken, H., Döring, U., Hardelauf, H., Jaekel, U., Hashagen, U., Neuendorf, O., Schwarze, H., Seidemann, R., 2000. Analysis of solute transport in a heterogeneous aquifer: the krauthausen field experiment. J. Contam. Hydrol. 45 (3), 329–358.

Voss, C., Souza, W.R., 1987. Variable density flow and solute transport simulation of regional aquifers containing a narrow freshwater-saltwater transition zone. Water Resour. Res. 23 (10), 1851–1866.

Ward, J.D., Simmons, C.T., Dillon, P.J., 2007. A theoretical analysis of mixed convection in aquifer storage and recovery: how important are density effects? J. Hydrol. 343 (3), 169–186.

Werner, A.D., Bakker, M., Post, V.E., Vandenbohede, A., Lu, C., Ataie-Ashtiani, B., Simmons, C.T., Barry, D.A., 2013. Seawater intrusion processes, investigation and management: recent advances and future challenges. Adv. Water Resour. 51, 3–26.

Xu, T., Gómez-Hernández, J.J., 2015. Inverse sequential simulation: a new approach for the characterization of hydraulic conductivities demonstrated on a non-Gaussian field. Water Resour. Res. 51 (4), 2227–2242. http://dx.doi.org/10.1002/2014WR016320.

Xu, T., Gómez-Hernández, J.J., 2016. Characterization of non-Gaussian conductivities and porosities with hydraulic heads, solute concentrations, and water temperatures. Water Resour. Res. 52 (8), 6111–6136.

Xu, T., Gómez-Hernández, J.J., Zhou, H., Li, L., 2013. The power of transient piezometric head data in inverse modeling: an application of the localized normal-score enkf with covariance inflation in a heterogenous bimodal hydraulic conductivity field. Adv. Water Resour. 54, 100–118.

Yeh, W.W.-G., 1986. Review of parameter identification procedures in groundwater hydrology: the inverse problem. Water Resour. Res. 22 (2), 95–108.

Yoon, H., McKenna, S.A., 2012. Highly parameterized inverse estimation of hydraulic conductivity and porosity in a three-dimensional, heterogeneous transport experiment. Water Resour. Res. 48 (10). http://dx.doi.org/10.1029/2012WR012149. W10536.

Zhang, Z.S., Jafarpour, B., Li, L.L., 2014. Inference of permeability heterogeneity from joint inversion of transient flow and temperature data. Water Resour. Res. 50 (6), 4710–4725.

Zhou, H., Gómez-Hernández, J.J., Franssen, H.-J.H., Li, L., 2011. An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. Adv. Water Resour. 34 (7), 844–864.

Zhou, H., Gómez-Hernández, J.J., Li, L., 2014. Inverse methods in hydrogeology: evolution and recent trends. Adv. Water Resour. 63, 22–37.

Zuurbier, K.G., Zaadnoordijk, W.J., Stuyfzand, P.J., 2014. How multiple partially penetrating wells improve the freshwater recovery of coastal aquifer storage and recovery (ASR) systems: a field and modeling study. J. Hydrol. 509, 430–441.