ORIGINAL PAPER

Comparison of regularized ensemble Kalman filter and tempered ensemble transform particle filter for an elliptic inverse problem with uncertain boundary conditions

Svetlana Dubinkina[1] 🆔 · **Sangeetika Ruchi[1]**

Abstract

In this paper, we focus on parameter estimation for an elliptic inverse problem. We consider a 2D steady-state single-phase Darcy flow model, where permeability and boundary conditions are uncertain. Permeability is parameterized by the Karhunen-Loeve expansion and thus assumed to be Gaussian distributed. We employ two ensemble-based data assimilation methods: ensemble Kalman filter and ensemble transform particle filter. The formal one approximates mean and variance of a Gaussian probability function by means of an ensemble. The latter one transforms ensemble members to approximate any posterior probability function. Ensemble Kalman filter considered here is employed with regularization and localization—R(L)EnKF. Ensemble transform particle filter is also employed with a form of regularization called tempering and localization—T(L)ETPF. Regularization is required for highly non-linear problems, where prior is updated to posterior via a sequence of intermediate probability measures. Localization is required for small ensemble sizes to remove spurious correlations. We have shown that REnKF outperforms TETPF. We have shown that localization improves estimations of both REnKF and TETPF. In numerical experiments when uncertainty is only in permeability, TLETPF outperforms RLEnKF. When uncertainty is both in permeability and in boundary conditions, TLETPF outperforms RLEnKF only for a large ensemble size 1000. Furthermore, when uncertainty is both in permeability and in boundary conditions but we do not account for error in boundary conditions in data assimilation, RLEnKF outperforms TLETPF.

Keywords Parameter estimation · Ensemble approximation · Tempering · Ensemble transform particle filter · Regularization · Ensemble Kalman inversion

1 Introduction

Ensemble-based data assimilation deals with estimation of uncertain parameters and states of a model constrained by available observations using an ensemble. It is widely employed in many fields, for example meteorology [15] and reservoir engineering [2]. While in meteorology one is interested in estimation of uncertain initial conditions of a high-dimensional chaotic system, in reservoir engineering—of estimating high-dimensional uncertain parameters, of permeability for example, of a deterministic non-chaotic system.

However, a source of model error is not only in random coefficients of a PDE. In inverse problems, another source

of model error is model reduction for example, where a complex model is replaced by a simple one. It has been acknowledged, i.e. [12], that accounting for model error in data assimilation greatly improves parameter-state estimation. Recent advances in accounting for model error in ensemble-based data assimilation are extension of iterative ensemble Kalman filter to include additive model error [25], adding model error in the randomized maximum likelihood though to correctly sample the posterior without marginalization [7], nonadditive though Gaussian model error update for Bayesian inversion [6], and adaptation of machine learning technics for data assimilation [18]. By additive model error we mean that if G is an erroneous approximation of the true model G^{true}, then $G(\cdot) = G^{\text{true}}(\cdot) + q$ where q is model error.

However, most of these works have considered either additive model error or Gaussian model error, with the sole exception of [7] where Gaussian anamorphosis was used. However, for high-dimensional problems, finding

✉ Svetlana Dubinkina
  s.dubinkina@cwi.nl

[1] CWI, Science park 123, 1098XG, Amsterdam,
  The Netherlands

Springer

a transformation to multivariate Gaussian probability is computationally challenging. Though the assumption of additive Gaussian model error simplifies an optimization problem, model error is not limited to being additive nor Gaussian. Therefore, it is essential for a data assimilation method to account for model error in a most general way. A straightforward example of such a data assimilation method is Markov Chain Monte Carlo (MCMC). However, for high-dimensional systems, it is impractical.

An alternative to MCMC is particle filtering [1]. Particle filtering is based on proposing an ensemble from a prior that is not necessarily close to the target posterior and to correct for this mismatch by computing importance weights. The importance weights are defined as a function of ensemble estimations and available observations. The ensemble is then resampled according to the estimated posterior. Particle filtering in its original form worked only for low-dimensional problems. However, due to recent advances of employing localization [10, 13] it has been proven to strive for high-dimensional problems as well.

There are different approaches to resampling in particle filtering, but most of them are stochastic. An ensemble transform particle filter [27] employs *deterministic* resampling, which reduces sampling error and thus needs a smaller ensemble than a typical particle filter. It also has a localized version. In [28], we have employed the method to an inverse problem of uncertain permeability. We have shown that though localization makes the ensemble transform particle filter deteriorates a posterior estimation of the leading modes, it makes the method applicable to high-dimensional problems. In [29], instead of localization, we have implemented tempering to the ensemble transform particle filter (TETPF). We have shown that iterations based on temperatures [21, 23] handle notably strongly nonlinear cases and that TETPF is able to predict multimodal distributions for high-dimensional problems.

In this paper, we consider a steady-state single-phase Darcy flow model. This groundwater model was first used as benchmark for inverse modelling in [11]. It has been used as a test model for the identification of parameters, for example with iterative regularization methods [20], an ensemble Kalman approach [4], and in our previous work with particle filtering [28, 29]. In this paper in addition to uncertain log permeability defined as a Gaussian process, we assume an error in boundary conditions that is non-Gaussian distributed. We note that error in boundary conditions of the groundwater model gives a non-linear response; thus, model error is non-additive. We compare T(L)ETPF to a regularized ensemble Kalman filter [3].

Regularized ensemble Kalman filter is a robust ensemble-based data assimilation method, which assumes Gaussian probabilities. It solves an optimization problem for mean and approximates variance with an ensemble. It

has been employed for history matching applications for example in [17, 22]. It has been shown that ensemble Kalman filter is able to estimate skewed probabilities with frequent observations [8, 26]. It, however, fails to estimate multimodal probabilities, e.g. [29].

Our goal is to investigate whether pressure estimation is sensitive to uncertain boundary conditions, and to compare T(L)ETPF with R(L)EnKF (with and without localization). We note that we have assumptions about precise information of sources of errors and prior probability density function. This is an idealized setting that however allows to check sensitivity of a data assimilation method to uncertainty in boundary conditions, which is also relevant for practical applications.

1.1 Bayesian inference

Both T(L)ETPF and R(L)EnKF are based on Bayesian inference. Assume $\boldsymbol{u} \in \tilde{\mathcal{U}}$ and $\boldsymbol{q} \in \tilde{\mathcal{Q}}$ are two independent random variables. Later on, we will have different assumptions on $\boldsymbol{u}$ and $\boldsymbol{q}$. We denote by $g : \tilde{\mathcal{U}} \times \tilde{\mathcal{Q}} \to \mathcal{Y}$ the non-linear forward operator that arises from a model under consideration. In other words, g maps the space $\tilde{\mathcal{U}} \times \tilde{\mathcal{Q}}$ of uncertain quantities to the observation space $\mathcal{Y}$ defined in terms of observable quantities, which are related to the solution of the model as

$$y = g(u, q).$$

Assume that $\boldsymbol{y}_{\text{obs}} \in \mathcal{Y}$ is an observation of y. Then according to the Bayes' formula

$$\pi(\boldsymbol{u}, \boldsymbol{q} | \boldsymbol{y}_{\text{obs}}) \propto \pi(\boldsymbol{y}_{\text{obs}} | \boldsymbol{u}, \boldsymbol{q}) \pi(\boldsymbol{u}) \pi(\boldsymbol{q}) \tag{1.1}$$

up to a constant of normalization, where π is a probability density function.

For any smooth function $f : \tilde{\mathcal{U}} \to \tilde{\mathcal{U}}$, its expectation is defined as

$$\overline{f(\boldsymbol{u})} = \int d\boldsymbol{u} f(\boldsymbol{u}) \int d\boldsymbol{q} \pi(\boldsymbol{u}, \boldsymbol{q} | \boldsymbol{y}_{\text{obs}}). \tag{1.2}$$

It is common, e.g. [8], to express the joint probability density function as

$$\pi(\boldsymbol{y}_{\text{obs}} | \boldsymbol{u}, \boldsymbol{q}) = \int \pi(\boldsymbol{y}_{\text{obs}}, y | \boldsymbol{u}, \boldsymbol{q}) d\boldsymbol{y} = \int \pi(\boldsymbol{y}_{\text{obs}} | \boldsymbol{y}) \pi(\boldsymbol{y} | \boldsymbol{u}, \boldsymbol{q}) d\boldsymbol{y} \tag{1.3}$$

and $\pi(\boldsymbol{y} | \boldsymbol{q}) = \delta(\boldsymbol{y} - g(\boldsymbol{u}, \boldsymbol{q}))$, where the transition density is the Dirac delta function.

2 Tempered ensemble transform particle filter

The goal of the Bayesian approach is to compute the posterior given by Eqs. 1.1–1.3. Sequential Monte Carlo

(SMC) is an approximation of the Bayesian posterior. An SMC method creates a finite sample from a prior, that is easy to sample from, and corrects for the differences between the prior and the posterior by computing so-called importance weights. Finally, a resampling is performed according to those weights in order to create a new sample.

2.1 Importance weights

We consider discrete random variables and define $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_M\} \subset \tilde{\mathcal{U}}$, $\mathbf{u}_i \in \mathbb{R}^n$, and $\mathcal{Q} = \{\mathbf{q}_1, \ldots, \mathbf{q}_M\} \subset \tilde{\mathcal{Q}}$, $\mathbf{q}_i \in \mathbb{R}^m$. The model has unknown quantities $\mathbf{u}^{\text{true}} \in \mathbb{R}^n$ and $\mathbf{q}^{\text{true}} \in \mathbb{R}^m$ that we wish to estimate from noisy observations $\mathbf{y}_{\text{obs}} \in \mathbb{R}^\kappa$, where $\kappa < n$ and $\kappa < m$.

$$\mathbf{y}_{\text{obs}} := g(\mathbf{u}^{\text{true}}, \mathbf{q}^{\text{true}}) + \eta,$$

where $\eta \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ with $\mathbf{R}$ being a known covariance matrix of the observation noise. Then the conditional probability density function is

$$\pi(\mathbf{y}_{\text{obs}}|\mathbf{y}) \propto \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{y}_{\text{obs}})'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{y}_{\text{obs}})\right].$$

Then a discrete approximation to Eq. 1.3 is

$$\pi(\mathbf{y}_{\text{obs}}|\mathbf{u}, \mathbf{q}) \propto \exp\left[-\frac{1}{2}\left(g(\mathbf{u}, \mathbf{q}) - \mathbf{y}_{\text{obs}}\right)'\mathbf{R}^{-1}\left(g(\mathbf{u}, \mathbf{q}) - \mathbf{y}_{\text{obs}}\right)\right],$$

where $'$ denotes the transpose. We assume the priors $\pi(\mathbf{u})$ and $\pi(\mathbf{q})$ are uniform, then denoting $\mathbf{v} = [\mathbf{u}\ \mathbf{q}]'$ the expectation of a function f of $\mathbf{v}$ is

$$\overline{f(\mathbf{v})} \approx \sum_{i=1}^{M} f(\mathbf{v}_i)w_i.$$

Here the importance weights are

$$w_i = \frac{h(\mathbf{v}_i)}{\sum_{j=1}^{M} h(\mathbf{v}_j)}, \quad \text{where} \quad h(\mathbf{v}) = \exp\left[-\frac{1}{2}\left(g(\mathbf{u}, \mathbf{q}) - \mathbf{y}_{\text{obs}}\right)'\mathbf{R}^{-1}\left(g(\mathbf{u}, \mathbf{q}) - \mathbf{y}_{\text{obs}}\right)\right]. \tag{2.1}$$

2.2 Tempering

An SMC method suffers when the likelihood $h(\mathbf{v})$ (2.1) is peaked, which could be due to very accurate data, amount of data, or when the prior poorly approximates the posterior. A tempering iterative approach tackles this problem by introducing temperatures $0 = \phi_0 < \cdots < \phi_T = 1$ and corresponding bridging likelihoods $h(\mathbf{v})^{(\phi_t - \phi_{t-1})}$ for $t = 1, \ldots, T$. A tempering parameter ϕ_t is typically chosen based on effective ensemble size

$$\text{ESS}_t(\phi_t) := \frac{\left(\sum_{i=1}^{M} w_i\right)^2}{\sum_{i=1}^{M} w_i^2}, \tag{2.2}$$

such that ESS does not drop below a certain threshold $1 < M_{\text{thresh}} < M$. A bisection algorithm on the interval $(\phi_{t-1}, 1]$ can be used to solve (2.2) [4]. If $\text{ESS}_t(1) > M_{\text{thresh}}$, then we can simply set $\phi_t = 1$ as no further tempering is thus required.

2.3 Deterministic resampling

In order to avoid filter degeneracy, each tempering iteration t needs to be supplied with resampling. Resampling is typically performed by a stochastic approach, which introduces an additional error. In TETPF, a tempering iteration t is accompanied by a *deterministic* resampling based on optimal transportation. This resampling transforms

particles with weights defined in terms of bridging likelihood

$$w_i^{(t)} = \frac{h\left(\mathbf{v}_i^{(t)}\right)^{(\phi_t - \phi_{t-1})}}{\sum_{j=1}^{M} h\left(\mathbf{v}_j^{(t)}\right)^{(\phi_t - \phi_{t-1})}},$$

where $h(\mathbf{v})$ is from Eq.2.1, to particles with uniform weights $1/M$ by maximizing the correlation between the particles. Thus, the optimal transport $\mathbf{S}$ is an $M \times M$ matrix with s_{ij} that satisfy

$$s_{ij} \geq 0, \quad \sum_{i=1}^{M} s_{ij} = \frac{1}{M}, \quad \sum_{j=1}^{M} s_{ij} = \frac{h\left(\mathbf{v}_i^{(t)}\right)^{(\phi_t - \phi_{t-1})}}{\sum_{j=1}^{M} h\left(\mathbf{v}_j^{(t)}\right)^{(\phi_t - \phi_{t-1})}}, \tag{2.3}$$

and minimizes the cost function

$$\sum_{i,j=1}^{M} s_{ij}\left\|\mathbf{v}_i^{(t)} - \mathbf{v}_j^{(t)}\right\|^2. \tag{2.4}$$

This gives rise to a resampling with replacement and a stochastic transport matrix $\mathbf{S}$. In order to have a deterministic optimal transformation, the following proposal is adopted

$$\tilde{\mathbf{v}}_j = M \sum_{i=1}^{M} \mathbf{v}_i^{(t)} \tilde{s}_{ij} \quad \text{for} \quad j = 1, \ldots, M, \tag{2.5}$$

where $\tilde{s}_{ij}$ is a solution to the optimization problem (2.3)–(2.4). To solve the linear transport problem for multivariate variables (2.3)–2.4, we use $FastEMD$ algorithm of [24]. Its computational complexity is of order $M^2 \ln M$, and the algorithm is available as a $MATLAB$ and a $Python$ subroutine.

2.3.1 Localization

Ensemble Transform Particle Filter as any particle filter does not have assumption about the posterior. Therefore, it still demands a large ensemble. For high-dimensional problems, this is computationally unfeasible. Hence, one has to decrease the number of degrees of freedom, i.e. by distance-based localization of [27, 28] abbreviated here LETPF.

Assume we have a numerical grid of $N \times N$ size with grid cells denoted by X^l for $l = 1, \ldots, N^2$. Assume that the uncertain parameter $\mathbf{u}$ is not grid-based. We assume, however, that there exists a matrix $\mathcal{A}$ such that $\log(\mathbf{k}) = \mathcal{A}\mathbf{u}$ is grid-based, thus $\log(k^l) = \log[k(X^l)]$. Then for the local update of an uncertain parameter $\log(k^l)$, we introduce a diagonal matrix $\hat{\mathbf{C}}^l \in \mathbb{R}^{\kappa \times \kappa}$ in the observation space with an element

$$(\hat{C}^l)_{\ell,\ell} = \rho \left(\frac{||X^l - r^\ell||}{r^{\mathrm{loc}}} \right) \quad \text{for} \quad \ell = 1, \ldots \kappa. \quad (2.6)$$

Here r^ℓ denotes the location of the observation, r^{loc} is a localization radius, and $\rho(\cdot)$ is a taper function, such as Gaspari-Cohn function by [9]

$$\rho(r) = \begin{cases} 1 - \frac{5}{3}r^2 + \frac{5}{8}r^3 + \frac{1}{2}r^4 - \frac{1}{4}r^5, & 0 \le r \le 1, \\ -\frac{2}{3}r^{-1} + 4 - 5r + \frac{5}{3}r^2 + \frac{5}{8}r^3 - \frac{1}{2}r^4 + \frac{1}{12}r^5, & 1 \le r \le 2, \\ 0, & 2 \le r. \end{cases}$$

$$(2.7)$$

The localization radius r^{loc} is typically tuned by a trial-and-error approach in terms of estimation error.

LETPF modifies the likelihood (2.1) as follows:

$$h^l(\mathbf{v}) = \exp \left[-\frac{1}{2} \left(g(\mathbf{u}, \mathbf{q}) - \mathbf{y}_{\mathrm{obs}} \right)' (\hat{\mathbf{C}}^l \mathbf{R}^{-1}) \left(g(\mathbf{u}, \mathbf{q}) - \mathbf{y}_{\mathrm{obs}} \right) \right], \quad (2.8)$$

where $\hat{\mathbf{C}}^l$ is the diagonal matrix given by Eq. 2.6. Then the optimal transport $\mathbf{S}^l$ is an $M \times M$ matrix with entries s_{ij}^l that satisfy

$$s_{ij}^l \ge 0, \quad \sum_{i=1}^M s_{ij}^l = \frac{1}{M}, \quad \sum_{j=1}^M s_{ij}^l = \frac{h^l \left(\mathbf{v}_i^{(t)} \right)^{(\phi_t - \phi_{t-1})}}{\sum_{j=1}^M h^l \left(\mathbf{v}_j^{(t)} \right)^{(\phi_t - \phi_{t-1})}}, \quad (2.9)$$

and minimizes the cost function

$$\sum_{i,j=1}^M s_{ij}^l \left[\log \left(k_i^{l,(t)} \right) - \log \left(k_j^{l,(t)} \right) \right]^2. \quad (2.10)$$

The estimated parameter $\log(\tilde{k}^l)$ is given by

$$\log(\tilde{k}_j^l) = M \sum_{i=1}^M \tilde{s}_{ij}^l \log \left(k_i^{l,(t)} \right) \quad \text{for} \quad j = 1, \ldots, M, \quad (2.11)$$

where $\tilde{s}_{ij}^l$ is a solution to the optimization problem (2.9)–(2.10). We note that localization reduces LETPF to a univariate transport problem. Solving the optimal transport problem for univariate variables (2.9)–(2.10) is computationally less expensive than for multivariate variables (2.3)–(2.4), since the marginal computational cost is in sorting M numbers. We use an algorithm described in [27] to solve the univariate linear transport problem (2.9)–(2.10). LETPF, however, loses direct dependence on the parameters at other grid cells. Update of the uncertain grid-based parameters $\log(k^l)$ could be performed in parallel for each $l = 1, \ldots, N^2$. Then the estimated model parameter is $\tilde{\mathbf{u}} = \mathcal{A}^{-1} \log(\tilde{\mathbf{k}})$.

To estimate $\mathbf{q}$, we solve the optimal transport problem with s_{ij}^{G} that satisfy

$$s_{ij}^{\mathrm{G}} \ge 0, \quad \sum_{i=1}^M s_{ij}^{\mathrm{G}} = \frac{1}{M}, \quad \sum_{j=1}^M s_{ij}^{\mathrm{G}} = \frac{h \left(\left[\tilde{\mathbf{u}}_i \ \mathbf{q}_i^{(t)} \right]' \right)^{(\phi_t - \phi_{t-1})}}{\sum_{j=1}^M h \left(\left[\tilde{\mathbf{u}}_j \ \mathbf{q}_j^{(t)} \right]' \right)^{(\phi_t - \phi_{t-1})}}, \quad (2.12)$$

and minimize the cost function

$$\sum_{i,j=1}^M s_{ij}^{\mathrm{G}} \left(\left\| \tilde{\mathbf{u}}_i - \tilde{\mathbf{u}}_j \right\|^2 + \left\| \mathbf{q}_i^{(t)} - \mathbf{q}_j^{(t)} \right\|^2 \right). \quad (2.13)$$

The estimated parameter $\tilde{\mathbf{q}}$ is given by

$$\tilde{\mathbf{q}}_j = M \sum_{i=1}^M \tilde{s}_{ij}^{\mathrm{G}} \mathbf{q}_i^{(t)}, \quad j = 1, \ldots, M, \quad (2.14)$$

where $\tilde{s}_{ij}^{\mathrm{G}}$ is a solution to the optimization problem (2.12)–(2.13). Finally, we set $\tilde{\mathbf{v}} = [\tilde{\mathbf{u}} \ \tilde{\mathbf{q}}]'$.

2.4 Mutation

The advantage of deterministic resampling is that it reduces sampling noise. The disadvantage of deterministic resampling is that for a deterministic and non-chaotic system, the filter collapse is unavoidable unless particle mutation is introduced. The mutation is performed over an

index $1 < \tau < \tau_{\max}$ with prescribed $\tau_{\max}$. At the first inner iteration $\tau = 1$, we assign $\mathbf{v} = \tilde{\mathbf{v}}$.

We denote by v_i^ℓ a component of $\mathbf{v}_i$, where $1 \le \ell \le n + m$. If v_i^ℓ has a Gaussian prior, then we use the preconditioned Crank-Nicolson pcn-MCMC method from [14]

$$v_i^{\ell,\mathrm{prop}} = \sqrt{1 - \beta^2}\, v_i^\ell + \beta \xi_i \quad \text{for} \quad i = 1, \ldots, M, \quad (2.15)$$

where $\{\xi_i\}_{i=1}^M$ is from normal distribution. The parameter $\beta \in [0,\ 1)$ is a free parameter. For uniform prior $U[a,\ b]$, we use random walk

$$v_i^{\ell,\ \mathrm{prop}} = v_i^\ell + \xi_i \quad \text{for} \quad i = 1, \ldots, M, \quad (2.16)$$

where $\xi_i \sim U[a - b,\ b - a]$, and we project $v_i^{\ell,\ \mathrm{prop}}$ to $[a,\ b]$ when necessary. The proposal (2.15)–(2.16) is accepted

$$\mathbf{v} = \mathbf{v}^{\mathrm{prop}} \quad \text{with the probability} \quad \min\left\{1, \frac{h(\mathbf{v}^{\mathrm{prop}})^{\phi_{t+1}}}{h(\mathbf{v})^{\phi_{t+1}}}\right\}, \quad (2.17)$$

and the inner iteration τ is increased by one. We choose β based on acceptance rate being between 20 and 30% by the last iteration T. The mutation (2.15)–(2.17) is repeated until $\tau = \tau_{\max}$, then we assign $\mathbf{v}^{(t+1)} = \mathbf{v}$.

After that, next tempering iteration proceeds by computing the weights (2.1), new temperature ϕ based on Eqs. 2.2 $\mathrm{ESS} \ge M_{\mathrm{thresh}}$, performing deterministic resampling either by Eqs. 2.3–2.5 for the non-localized method or by Eqs. 2.9–2.14 for the localized method, and concluding by mutation (2.15)–(2.17) for $\tau_{\max}$ iterations. The algorithms stops when the final temperature ϕ reaches one. Recall that T is the total number of tempering iterations that corresponds to ϕ reaching one. Thus, T is not predefined, which can lead to computationally unfeasible iteration times. TETPF demands $TM(\tau_{\max} + 1)$ evaluations of the model g, and TLETPF demands $TM(\tau_{\max} + 2)$ evaluations of the model g.

3 Regularized ensemble Kalman Filter

REnKF is based on the Ensemble Kalman Filter with perturbed observations

$$\mathbf{y}_i^\eta = \mathbf{y}_{\mathrm{obs}} + \boldsymbol{\eta}_i \quad \text{for} \quad i = 1, \ldots, M,$$

where $\boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ with $\mathbf{R}$ being a known covariance matrix of the observation noise. We define an M-dimensional vector with all elements equal to 1 as $\mathbf{1}_M$. REnKF solves the following set of equations for $t = 0, \ldots, T - 1$ with $\mathbf{v}^{(0)}$ being an initial ensemble

$$\mathbf{B}^{\mathrm{gg}} = \frac{1}{M-1}\left(g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)}) - \overline{g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)})}\mathbf{1}_M'\right)\left(g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)}) - \overline{g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)})}\mathbf{1}_M'\right)',$$

$$\mathbf{B}^{\mathrm{vg}} = \frac{1}{M-1}\left(\mathbf{v}^{(t)} - \overline{\mathbf{v}^{(t)}}\mathbf{1}_M'\right)\left(g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)}) - \overline{g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)})}\mathbf{1}_M'\right)',$$

$$\mathbf{v}_i^{(t+1)} = \mathbf{v}_i^{(t)} + \mathbf{B}^{\mathrm{vg}}\left(\mathbf{B}^{\mathrm{gg}} + \mu^{(t)}\mathbf{R}\right)^{-1}\left(\mathbf{y}_i^\eta - g(\mathbf{u}_i^{(t)}, \mathbf{q}_i^{(t)})\right) \quad \text{for} \quad i = 1, \ldots, M. \quad (3.1)$$

The regularized parameter $\mu^{(t)}$ is chosen such that

$$\mu^{(t)}\left\| \mathbf{R}^{1/2}\left(\mathbf{B}^{\mathrm{gg}} + \mu^{(t)}\mathbf{R}\right)^{-1}\left(\mathbf{y}_{\mathrm{obs}} - \overline{g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)})}\right)\right\|$$
$$\ge \Omega\left\| \mathbf{R}^{-1/2}\left(\mathbf{y}_{\mathrm{obs}} - \overline{g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)})}\right)\right\| \quad (3.2)$$

for predefined $\Omega \in (0, 1)$. This is achieved by the bisection method $\mu^{\tau+1} = 2^\tau \mu^0$ for $\tau = 0, \ldots, \tau_{\max}$ and an initial guess μ^0. We assign $\mu^{(t)} = \mu^{\tau_{\max}}$, where $\tau_{\max}$ is the first integer for which (3.2) holds.

Finally, REnKF is stopped based on discrepancy principle, namely when

$$\left\| \mathbf{R}^{-1/2}\left(\mathbf{y}_{\mathrm{obs}} - \overline{g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)})}\right)\right\| \le 1/\Omega\left\| \mathbf{R}^{-1/2}\eta\right\| \quad (3.3)$$

with η being the observation noise. Thus, the total number of iterations T is not predefined, as in T(L)ETPF. The rule of thumb is to choose $\Omega \in (0.5, 1)$, and we choose $\Omega = 0.7$ for all the numerical experiments. REnKF demands $TM + 1$ evaluations of the model g.

3.1 Localization

In Ensemble Kalman filter, covariance-based localization of the Kalman gain [16, 19] can be applied in order to remove spurious correlations due to a small ensemble size. We assume again having a numerical grid of $N \times N$ size with grid cells denoted by X^l for $l = 1, \ldots, N^2$. Assume that the uncertain parameter $\mathbf{u}$ is not grid-based. We assume, however, that there exists a matrix $\mathcal{A}$ such that $\log(\mathbf{k}) = \mathcal{A}\mathbf{u}$ is grid-based, thus $\log(k^l) = \log[k(X^l)]$. Then Eq. 3.1 for a localized EnKF, denoted here LEnKF, is rewritten as

$$\log\left(\mathbf{k}_i^{(t+1)}\right) = \log\left(\mathbf{k}_i^{(t)}\right) + \hat{\mathcal{C}} \circ \mathbf{B}^{\mathrm{kg}}\left(\mathbf{B}^{\mathrm{gg}} + \mu^{(t)}\mathbf{R}\right)^{-1}$$
$$\times \left(\mathbf{y}_i^\eta - g(\mathbf{u}_i^{(t)}, \mathbf{q}_i^{(t)})\right) \quad \text{for} \quad i = 1, \ldots, M,$$

$$\mathbf{q}_i^{(t+1)} = \mathbf{q}_i^{(t)} + \mathbf{B}^{\mathrm{q\,g}}\left(\mathbf{B}^{\mathrm{gg}} + \mu^{(t)}\mathbf{R}\right)^{-1}$$
$$\times \left(\mathbf{y}_i^\eta - g(\mathbf{u}_i^{(t)}, \mathbf{q}_i^{(t)})\right) \quad \text{for} \quad i = 1, \ldots, M.$$

Here ∘ denotes the element-wise product and $\hat{\mathcal{C}}$ is a distance-based correlation matrix, an element of which is

$$\hat{\mathcal{C}}_{l,\ell} = \rho\left(\frac{||X^l - r^\ell||}{r^{\text{loc}}}\right) \quad \text{for } l = 1, \dots, N^2 \text{ and } \ell = 1, \dots, \kappa, \tag{3.4}$$

where r^ℓ denotes the location of the observation, r^{loc} is a localization radius, and ρ is given by Eq. 2.7. The covariance matrices $\mathbf{B}^{gg}$, $\mathbf{B}^{kg}$ and $\mathbf{B}^{qg}$ are as follow:

$$\mathbf{B}^{gg} = \frac{1}{M-1}\left(g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)}) - \overline{g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)})}\mathbf{1}'_M\right)$$
$$\times \left(g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)}) - \overline{g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)})}\mathbf{1}'_M\right)',$$

$$\mathbf{B}^{kg} = \frac{1}{M-1}\left(\log(\mathbf{k}^{(t)}) - \overline{\log(\mathbf{k}^{(t)})}\mathbf{1}'_M\right)$$
$$\times \left(g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)}) - \overline{g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)})}\mathbf{1}'_M\right)',$$

$$\mathbf{B}^{qg} = \frac{1}{M-1}\left(\mathbf{q}^{(t)} - \overline{\mathbf{q}^{(t)}}\mathbf{1}'_M\right)\left(g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)}) - \overline{g(\mathbf{u}^{(t)}, \mathbf{q}^{(t)})}\mathbf{1}'_M\right)'.$$

RLEnKF also demands $TM + 1$ evaluations of the model g, as REnKF.

4 Numerical experiment

We consider a test case of estimating uncertain Gaussian permeability. We consider a steady-state single-phase Darcy flow model defined over an aquifer of two-dimensional physical domain $D = [0, 6] \times [0, 6]$, which is given by

$$-\nabla \cdot [k(x, y)\nabla P(x, y)] = \mathcal{F}(x, y) \quad \text{for } (x, y) \in D,$$

where $\nabla = (\partial/\partial x \ \partial/\partial y)'$, $\cdot$ is the dot product, $P(x, y)$ is the pressure, $k(x, y)$ is the permeability, and the source term is

$$\mathcal{F}(x, y) = \begin{cases} 0 & \text{for } 0 \le y \le 4, \\ 137 & \text{for } 4 < y \le 5, \\ 274 & \text{for } 5 < y \le 6. \end{cases}$$

The boundary conditions are a combination of Dirichlet and Neumann boundary conditions

$$P(x, 0) = 100, \quad \frac{\partial P}{\partial x}(6, y) = 0,$$
$$-k(0, y)\frac{\partial P}{\partial x}(0, y) = 500(1 + q), \quad \frac{\partial P}{\partial y}(x, 6) = 0,$$

where q denotes error in boundary conditions.

We implement a cell-centred finite difference method to discretize the domain D into $N \times N$ grid cells X^l of size Δx^2. We solve the true forward model G on a fine grid $N = N_{\text{f}} = 140$ for the true solution. Then the synthetic observations are obtained by

$$\mathbf{y}_{\text{obs}} = \mathbf{L}(\mathbf{P}^{\text{true}}) + \boldsymbol{\eta}.$$

An element of $\mathbf{L}(\mathbf{P}^{\text{true}})$ is a linear functional of pressure, namely

$$L^\ell(\mathbf{P}^{\text{true}}) = \frac{1}{2\pi\sigma^2}\sum_{l=1}^{N_{\text{f}}^2} \exp\left(-\frac{||X^l - r^\ell||^2}{2\sigma^2}\right)$$
$$\mathbf{P}^{\text{true},l}\Delta x_{\text{f}}^2 \quad \text{for } \ell = 1, \dots, \kappa,$$

where r^ℓ denotes the location of the observation, $\kappa = 36$ the number of observations, $\Delta x_{\text{f}} = 6/N_{\text{f}}$, and $\sigma = 0.01$. The observation locations are displayed in Fig. 4 as circles.

Observation noise is denoted by $\boldsymbol{\eta}$ and it is drawn from $\mathcal{N}(\mathbf{0}, \mathbf{R})$. Observation error covariance $\mathbf{R}$ is known, and we choose it such that the norm of the noise is 1% of the norm of the data. Such a small noise makes the data assimilation problem hard to solve, since the likelihood is very peaked and a non-iterative data assimilation approach fails.

Both the true permeability and an initial ensemble are drawn from the same prior distribution as the prior includes knowledge about geological properties. We assume log permeability is generated by a random draw from a Gaussian distribution $\mathcal{N}(\log(\mathbf{5}), \mathbf{C})$. Here $\mathbf{5}$ is an N^2 vector with all elements being 5 and $\mathbf{C}$ is Whittle-Matern correlation, an element of which is given by

$$C^{l\ell} = \frac{1}{\Gamma(1)}\frac{d^{l\ell}}{\delta}\Upsilon_1\left(\frac{d^{l\ell}}{\delta}\right) \quad \text{for } l, \ell = 1, \dots, N^2.$$

Here $d^{l\ell}$ is the distance between two spatial locations, $\delta = 0.5$ is the correlation length, Γ is the gamma function, and Υ_1 is the modified Bessel function of the second kind of order one. We denote by λ and γ eigenvalues and eigenfunctions of $\mathbf{C}$, respectively, then following Karhunen-Loeve expansion log permeability is

$$\log(k^l) = \log(5) + \sum_{\ell=1}^{N^2}\sqrt{\lambda^\ell}\gamma^{\ell l}u^\ell \quad \text{for } l = 1, \dots, N^2,$$

where u^ℓ is i.i.d. from a normal distribution for $\ell = 1, \dots, N^2$.

Therefore, the initial parameter u is drawn from $\mathcal{N}(0, 1)$, while the initial boundary condition error q is drawn from a uniform distribution $U[0 \ 0.5]$. We then solve the incorrect forward model g on a coarse grid $N = N_{\text{c}} = 70$. The uncertain parameter $\mathbf{u}$ has the dimension $n = 4900$, which makes the dimension of $\mathbf{v}$ $n + m = 4901$. We perform 40 numerical experiments with both T(L)ETPF and R(L)EnKF to check initial sample sensitivity. We conduct numerical experiments with ensemble sizes 100 and 1000. We compare the methods with a pcn-MCMC method without model error in boundary conditions. An MCMC experiment was conducted using 100 chains with the lengths 10^6, burn-in period 10^5, and thinning period 10^3

each. For T(L)ETPF, we choose $\tau_{\max} = 20$ and $\beta = 0.045$ for mutation, since it gives good mixing with acceptance rate at the final tempering iteration around 20%. We set the threshold for ESS to be $M_{\text{thresh}} = M/3$.

We define the root mean square error (RMSE) of a mean field $\overline{\Xi} = 1/M \sum_{i=1}^{M} \Xi_i$ as

$$\text{RMSE}(\Xi) = \sqrt{\left(\overline{\Xi} - \Xi^{\text{MCMC}}\right)^T \left(\overline{\Xi} - \Xi^{\text{MCMC}}\right)}$$

for either log permeability $\Xi = \log(\mathbf{k})$ or pressure $\Xi = \mathbf{P}$. To choose a favouring localization radius, we perform a numerical experiment with r^{loc} ranging from one to six with an increment of one. Then we define the favouring localization radius as a localization radius that gives the smallest RMSE in terms of mean log permeability for that numerical experiment. For TLETPF, the favouring localization radius is $r^{\text{loc}} = 1$ for both ensemble sizes 100 and 1000. For RLEnKF, the favouring localization radius is $r^{\text{loc}} = 3$ for both ensemble sizes 100 and 1000.

4.1 Data assimilation without localization

In this section, we compare REnKF to TETPF, thus without localization. In Fig. 1, we plot RMSE of mean log permeability on the left and of mean pressure on the right at different ensemble sizes for both TETPF and REnKF. When uncertainty is only in permeability, REnKF outperforms TETPF for estimation of $\log(k)$ and as a consequence for estimation of pressure, as it can be seen in Fig. 1 a and b and as it has been reported in the literature, e.g. [28, 29]. When uncertainty is in both permeability and boundary conditions, we investigate methods performance for two numerical set-ups. First numerical set-up is when we account for model error in boundary conditions. Second numerical set-up is when we do not account for model error in boundary conditions. In both set-ups, REnKF outperforms TETPF as seen in Fig. 1c–f.

Let us now examine numerical experiments in more detail. First, we compare numerical experiments when we account for error in boundary conditions to numerical experiments with no error in boundary conditions. Comparing Fig. 1d–b, we observe that pressure estimation does not change. At the same time comparing Fig. 1c–a, we observe that log permeability estimation becomes worse when model error is present, as to be expected.

Next, we compare numerical experiments when we do not account for error in boundary conditions with numerical experiments when we do account for error in boundary conditions. Comparing Fig. 1e–c, we observe that for REnKF at ensemble size 1000 and TETPF at both ensemble sizes, log permeability is better estimated when a method does not account for error in boundary conditions. Pressure

is also better estimated by both methods when error in boundary conditions is not accounted for, as seen from comparing Fig. 1f–d.

In order to further investigate this result, we compare REnKF and TETPF with an estimation from a pcn-MCMC with error in boundary conditions. In Fig. 2, we plot the posterior approximations of q. We recall that the prior for q is $U[0\ 0.5]$. We observe that MCMC gives a skewed posterior. TETPF has better resemblance to the posterior obtained by MCMC, while REnKF gives negative values of q (negative values of q for TETPF and MCMC are only due to kernel representation for plotting). We, however, observed that pressure is better estimated by both methods when error in boundary conditions is not accounted for. Thus, by allowing for error in boundary conditions and not constraining it by data, we obtain better pressure estimation.

4.2 Data assimilation with localization

Next we compare RLEnKF with TLETPF, thus with localization. In Fig. 3, we plot RMSE of mean log permeability on the left and of mean pressure on the right at different ensemble sizes for both TLETPF and RLEnKF. When uncertainty is only in permeability, TLETPF outperforms RLEnKF for estimation of $\log(k)$ and as a consequence for estimation of pressure, as it can be seen in Fig. 3 a and b. This shows that localization drastically improves TETPF, such that it outperforms RLEnKF even for normally distributed $\log(k)$.

Next, we compare numerical experiments when we account for error in boundary conditions with numerical experiments with no error in boundary conditions. Comparing Fig. 3d–b, we observe that pressure estimation does not change. At the same time comparing Fig. 3c–a, we observe that log permeability estimation becomes worse when model error is present, as to be expected. Here TLETPF again outperforms RLEnKF but only at a large ensemble size 1000. For comparison, we show results of MCMC, when error in boundary conditions is accounted for. In Fig. 4, we plot mean of log permeability at the top and variance of log permeability at the bottom for MCMC, TLETPF, and RLEnKF at ensemble size 1000. We observe that both methods give a reasonably good approximation of the MCMC mean log permeability. The variance, however, is underestimated.

Next, we compare numerical experiments when we do not account for error in boundary conditions with numerical experiments when we do account for error in boundary conditions. Comparing Fig. 3e–c, we observe that for RLEnKF and TLETPF log permeability is better estimated when a method accounts for error in boundary conditions. This is opposite to results of REnKF and TETPF. When it comes to pressure estimation, TLETPF gives equivalent
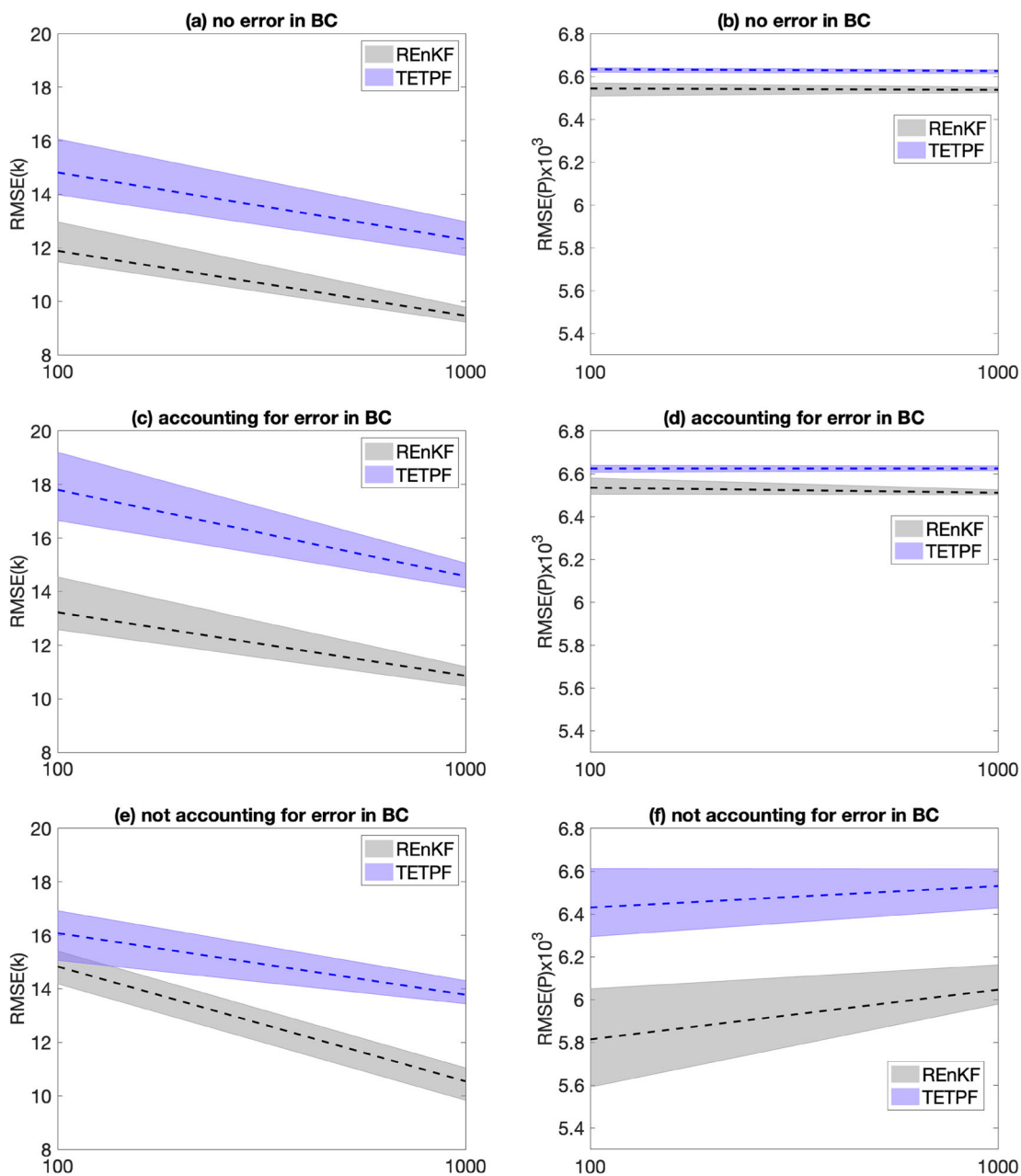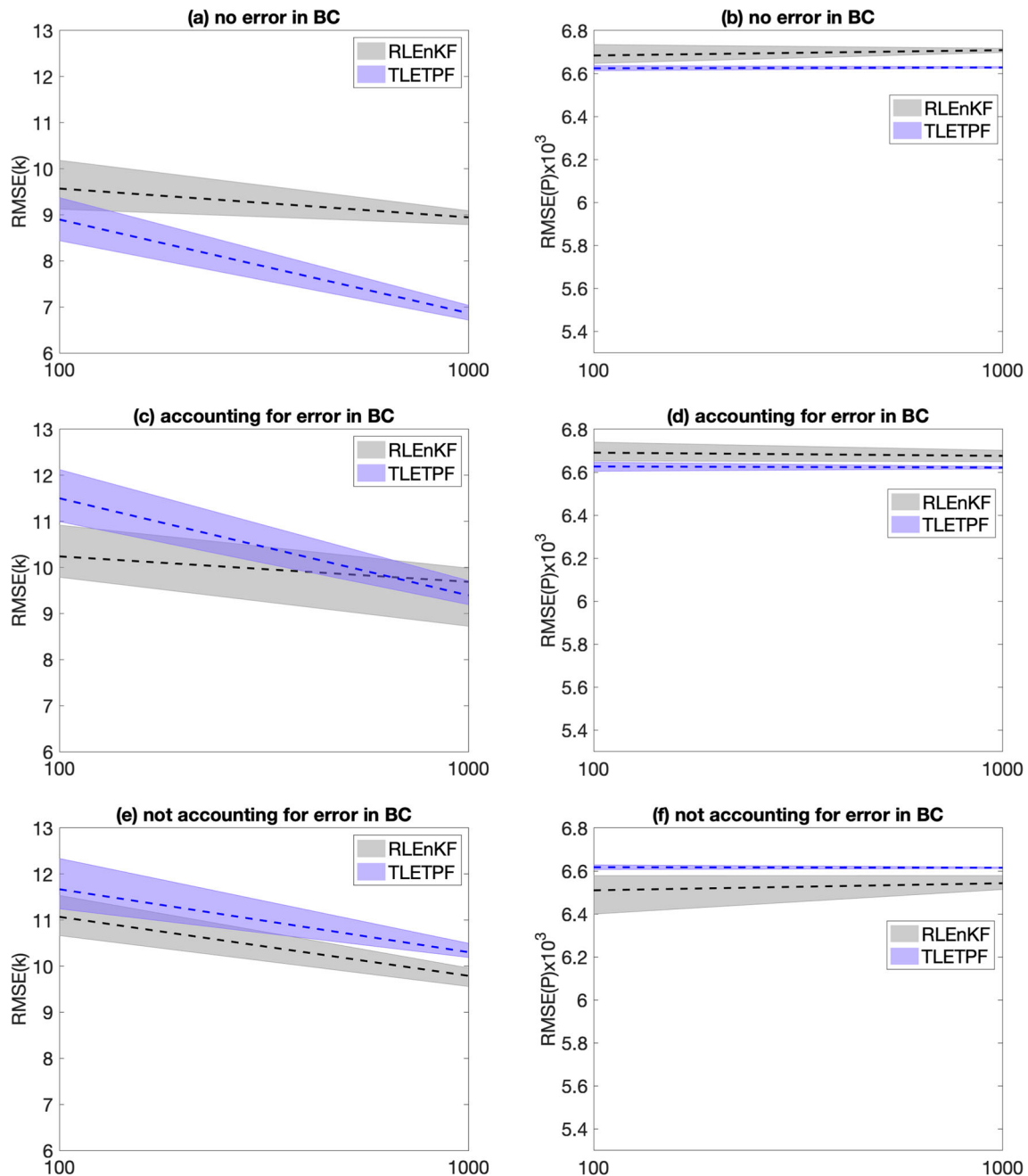
Fig. 1 RMSE for log(**k**) is on the left. RMSE $\times 10^3$ for pressure is on the right. A dashed line is for the median, a shaded area is for 25 and 75 percentiles over 40 simulations. Ensemble size is on *x*-axis. REnKF is shown in grey. TETPF is shown in blue. **a**, **b** For no model error, **c**, **d** for accounting for model error, and **e**, **f** for not accounting for model error

Fig. 2 Posterior of error in boundary conditions q. On the left is TETPF and on the right is REnKF. Results for the ensemble size 100 are shown in blue and for 1000 in pink, where one line is for one simulation out of 40. MCMC is shown in black
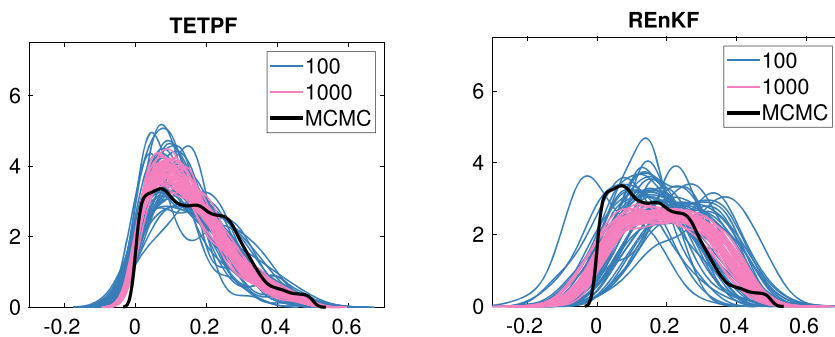
Fig. 3 RMSE for log(**k**) is on the left. RMSE $\times 10^3$ for pressure is on the right. A dashed line is for the median, a shaded area is for 25 and 75 percentiles over 40 simulations. Ensemble size is on *x*-axis. RLEnKF is shown in grey. TLETPF is shown in blue. **a**, **b** for no model error, **c**, **d** for accounting for model error, and **e**, **f** for not accounting for model error

performance for all three test cases. This means that TLETPF is not sensitive to error in boundary conditions in terms of pressure estimation. RLEnKF, however, better estimates pressure when error in boundary conditions is not accounted for, as seen from comparing Fig. 3f–d. The same was observed in experiments with REnKF. When comparing RLEnKF and TLETPF in terms of the posterior approximations of *q* shown in Fig. 5, we observe that none of the methods gives estimations close to MCMC.

4.2.1 R(L)EnKF at large ensemble size

We further investigate R(L)EnKF (with and without localization) performance in case when error in boundary conditions is not accounted for. We increase ensemble size to $M = 6000$ for localized version and to $M = 7700$ for non-localized version. We have tested localization radius between 1 and 8 with an increment of 1, and $r^{loc} = 6$ gave the smallest RMSE. We compute RMSE of mean log

Fig. 4 Mean log(**k**) at the top and variance of log(**k**) at the bottom for a numerical experiment where error in boundary conditions is accounted for. MCMC is on the left, with circles for the observation locations. A simulation with ensemble size 1000 is in the middle for TLETPF and on the right for RLEnKF.
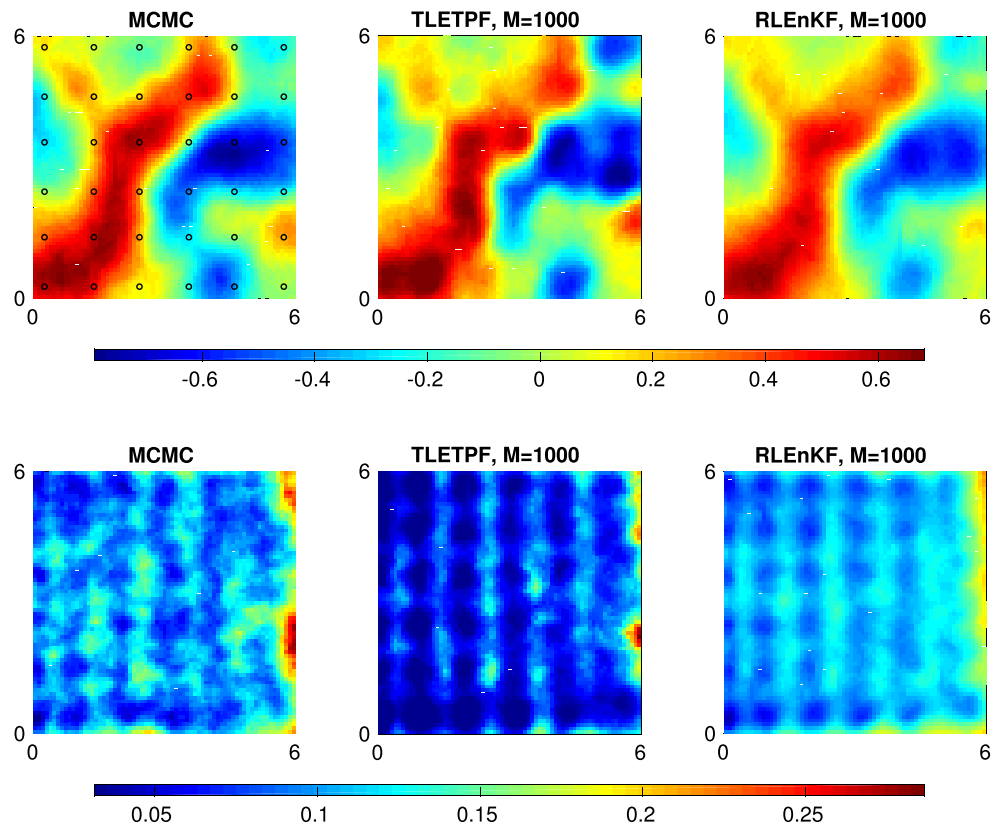
Fig. 5 Posterior of error in boundary conditions q. On the left is TLETPF and on the right is RLEnKF. Results for the ensemble size 100 are shown in blue and for 1000 in pink, where one line is for one simulation out of 40. MCMC is shown in black
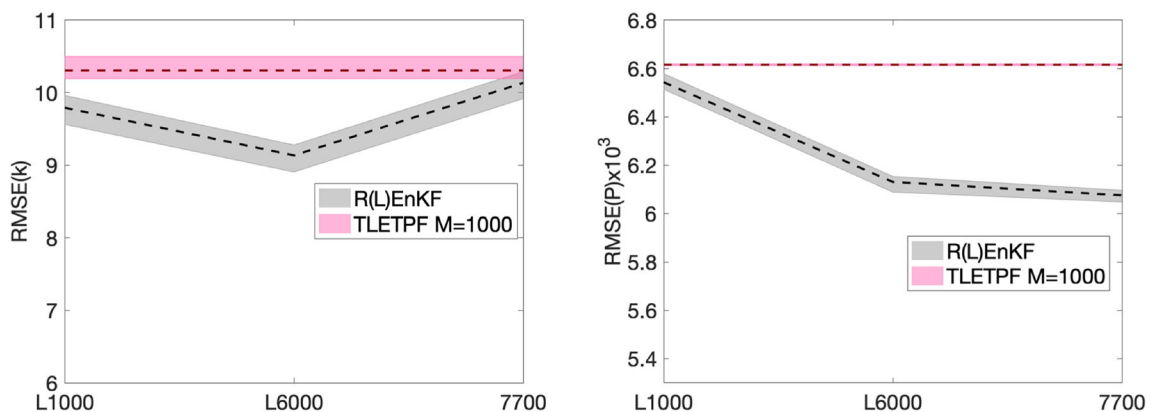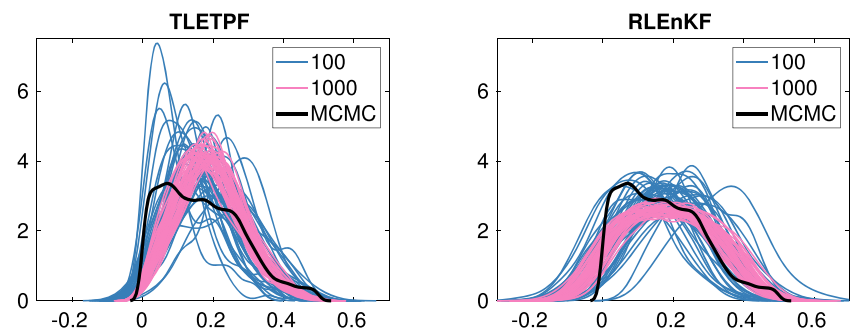
Fig. 6 RMSE for log(**k**) is on the left, and RMSE $\times 10^3$ for pressure is on the right for the test case where error in boundary conditions is not accounted for. A dashed line is for the median, a shaded area is for 25 and 75 percentiles over 40 simulations. On the x-axis, numbers stand for ensemble sizes and L stands for a localized method. R(L)EnKF is shown in grey. TLETPF for ensemble size 1000 is shown in pink

Table 1 Mean over 40 simulations of number of iterations for T(L)ETPF and R(L)EnKF at ensemble size 1000

	TETPF	TLETPF	REnKF	RLEnKF
No ME	13	28	8	11
Accounting for ME	14	35	11	15
Not accounting for ME	19	30	19	19

permeability and mean pressure and display them in Fig. 6 in grey for R(L)EnKF. We observe that as ensemble size increases, pressure becomes better estimated. For 1000 and 6000 members, the improvement in P can be explained by the improvement in $\log(k)$, as seen on the left of Fig. 6. Comparing RLEnKF with 6000 members to REnKF with 7700 members, we observe that better pressure estimation can only be explained by better estimation of q. In Fig. 6, we show TLETPF with ensemble size 1000 in pink. We observe that R(L)EnKF outperforms TLETPF greatly. For TLETPF, we were not able to increase the ensemble size beyond 1000 due to high computational costs.

4.3 Computational costs

In Table 1, we show the number of iterations a method takes on average. We should note that for R(L)EnKF, we put a limit of 20 iterations. When performing numerical experiments where no model error is present or when model error is accounted for, we observed that R(L)EnKF attains the stopping criterion (3.3) in less than 20 iterations for $\Omega = 0.7$. When performing numerical experiments when model error is not accounted for, we observed that in the majority of experiments after 20 iterations, the data misfit starts to increase and the stopping criterion is not attained. Decreasing Ω to values less than 0.5 solves the issue of attaining the stopping criterion but the data misfit is too high for sensible estimations. By the try-and-error approach, we discovered that it is preferable to keep $\Omega = 0.7$ and put a limit of 20 total iterations. For T(L)ETPF, we do not put any limit on total number of iterations. The wall clock of one iteration of TETPF is 70 s, of TLETPF is 78 s, of REnKF is 9 s, and of RLEnKF is 10 s.

5 Conclusions

We have considered 2D steady-state Darcy flow with uncertain permeability. Observations of pressure at 36 locations were used to infer permeability at a grid 70×70. The corresponding inverse problem was solved using two methods: regularized (localized) ensemble Kalman filter—R(L)EnKF—and tempered (localized) ensemble transform particle filter—T(L)ETPF. Ensemble Kalman filter updates uncertain parameters based on an assumption of Gaussian probabilities. Therefore, mean and variance need to be estimated. Ensemble Kalman filter makes further assumption that variance can be approximated by an ensemble. Due to small ensemble size, variables exhibit spurious correlations. Spurious correlations can be removed by employing localization—where covariance matrix in Kalman gain is multiplied by a distance-based matrix, for example. Moreover, it has been recognized that ensemble Kalman filter overfits observations and thus requires regularization. Therefore, in our numerical experiments, we have used regularized (localized) ensemble Kalman filter.

Ensemble transform particle filter makes an assumption that probability can be approximated by empirical measure. It employs optimal transport to transform prior measure into posterior. It correctly estimates mean but not variance. Ensemble transform particle filter is computationally demanding algorithm, since its complexity is of order $M^2 \ln M$ where M is ensemble size. Ensemble transform particle filter suffers spurious correlations as ensemble Kalman filter and thus requires localization. There exists only one type of localization for ensemble transform particle filter, which requires solving a univariate optimal transport problem at every grid though that can be done in parallel. It has been also recognized that when observations are very accurate or numerous, ensemble transform particle filter needs tempering—where prior is transformed into posterior over iterations. Therefore, in our numerical experiments, we have used tempered (localized) ensemble transform particle filter.

We have performed experiments with ensemble sizes 100 and 1000 for both R(L)EnKF and T(L)ETPF. We have shown that localization improves estimations obtained by REnKF and TETPF in all considered numerical experiments with 2D steady-state Darcy flow. First, we have considered uncertainty only in permeability. For ensemble sizes 100 and 1000, we have shown that TLETPF outperforms both REnKF and RLEnKF. Next, we have considered uncertain permeability and uncertain boundary conditions. We have shown that TLETPF outperforms both REnKF and RLEnKF but only at a large ensemble size of 1000. Finally, we have considered uncertainty in both permeability and boundary conditions but we did not account for error in boundary conditions in data assimilation. In this set-up, RLEnKF outperforms both TETPF and TLETPF.

T(L)ETPF is computationally more demanding than R(L)EnKF and requires more iterations than R(L)EnKF. Moreover, TETPF used here is first-order accurate, since it underestimates the ensemble spread. There exists second-order accurate ensemble transform particle filter [5], which we plan to investigate for an inverse problem in future.

References

1. Doucet, A., de Freitas, N., Gordon, N.: Sequential Monte-Carlo Methods in Practice. Springer, New York (2001)
2. Emerick, A.A., Reynolds, A.C.: Ensemble smoother with multiple data assimilation. Comput. Geosci. **55**, 3–15 (2013)
3. Iglesias, M.A.: Iterative regularization for ensemble data assimilation in reservoir models. Comput. Geosci. **19**(1), 177–212 (2015)
4. Iglesias, M.A.: A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems. Inverse Probl. **32**, 025002 (2016)
5. Acevedo, W., de Wiljes, J., Reich, S.: Second-order accurate ensemble transform particle filters. SIAM J. Sci. Comput. **39**(5), A1834–A1850 (2017)
6. Daniela, C., Dunlop, M., Somersalo, E., Stuart A.: Iterative updating of model error for Bayesian inversion. Inverse Probl. **34**(2), 025008 (2018)
7. Oliver, D.: Metropolized randomized maximum likelihood for improved sampling from multimodal distributions. SIAM/ASA J. Uncertain. Quant. **5**(1), 259–277 (2017)
8. Evensen, G.: Analysis of iterative ensemble smoothers for solving inverse problems. Comput. Geosci. **22**(3), 885–908 (2018)
9. Gaspari, G., Cohn, S.E.: Construction of correlation functions in two and three dimensions. Q. J. Roy. Meteorol. Soc. **125**(554), 723–757 (1999)
10. Penny, S.G.: Miyoshi T. A local particle filter for high-dimensional geophysical systems. Nonlin. Processes Geophys. **23**, 391–405 (2016)
11. Carrera, J., Neuman, S.P.: Estimation of aquifer parameters under transient and steady state conditions: III. Application to synthetic and field data. Water Resour. Res., 22 (1986)
12. Kaipio, J., Somersalo, E.: Statistical and Computational Inverse Problems. Springer, New York (2004)
13. Poterjoy, J.A.: Localized particle filter for high-dimensional nonlinear systems. Mon. Wea. Rev. **144**(1), 59–76 (2016)
14. Cotter, S.L., Roberts, G.O., Stuart, A.M., White, D.: MCMC methods for functions: modifying old algorithms to make them faster. Stat. Sci. **28**, 424–446 (2013)
15. Houtekamer, P.L., Lefaivre, L., Derome, J., Ritchie, H., Mitchell, H.L.: A system simulation approach to ensemble prediction. Mon. Wea. Rev. **124**, 1225–1242 (1996)
16. Houtekamer, P.L., Mitchell H. L.: A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Wea. Rev. **129**, 123–137 (2001)
17. Luo, X., Stordal, A.S., Lorentzen, R.J., Naevdal, G.: Iterative ensemble smoother as an approximate solution to a regularized minimum-average-cost problem: Theory and applications. Soc. Petrol. Eng. J. **20**(5), 962–982, 10 (2015)
18. Luo, X.: Ensemble-based kernel learning for a class of data assimilation problems with imperfect forward simulators. PLOS ONE **14**(7), 1–40, 07 (2019)
19. Hamill, T.M., Whitaker, J.S.: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman. Filter. Mon. Wea Rev. **129**, 2776–2790 (2001)
20. Hanke, M.: A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems. Inverse Probl. **13**, 79–95 (1997)
21. Neal, R.M.: Annealed importance sampling. Stat. Comput. **11**(2), 125–139 (2001)
22. Ma, X., Bi, L.: A robust adaptive iterative ensemble smoother scheme for practical history matching applications. Comput. Geosci. **23**(3), 415–442 (2019)
23. Kantas, N., Beskos, A., Jasra, A.: Sequential Monte Carlo for inverse problems: a case study for the Navier Stokes equation. SIAM/ASA J. Uncertain. Quant. **2**, 464–489 (2014)
24. Pele, O., Werman, M.: Fast and robust earth mover's distances. In: IEEE 12th international conference on computer vision, pp. 460–467. IEEE (2009)
25. Sakov, P., Haussaire, J.-M., Bocquet, M.: An iterative ensemble Kalman filter in the presence of additive model error. Q. J. Roy. Meteorol. Soc. **144**(713), 1297–1309 (2018)
26. Dubinkina, S.: Relevance of conservative numerical schemes for an ensemble Kalman filter. Q. J. Roy. Meteorol. Soc. **144**(711), 468–477 (2018)
27. Reich, S., Cotter, C.: Probabilistic Forecasting and Bayesian Data Assimilation. Cambridge University Press, Cambridge (2015)
28. Ruchi, S., Dubinkina, S.: Application of ensemble transform data assimilation methods for parameter estimation in reservoir modeling Nonlin. Process. Geophys. **25**, 731–746 (2018)
29. Ruchi, S., Dubinkina, S., Iglesias, M.A.: Transform-based particle filtering for elliptic Bayesian inverse problems (2018)