

Transform-based particle filtering for elliptic Bayesian inverse problems

S Ruchi¹, S Dubinkina¹ and M A Iglesias²

¹ Centrum Wiskunde & Informatica, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

² School of Mathematical Sciences, The University of Nottingham, University Park, Nottingham, NG7 2RD, UK

E-mail: s.dubinkina@cwi.nl

August 2017

Abstract. We introduce optimal transport based resampling in adaptive SMC. We consider elliptic inverse problems of inferring hydraulic conductivity from pressure measurements. We consider two parametrizations of hydraulic conductivity: by Gaussian random field, and by a set of scalar (non-)Gaussian distributed parameters and Gaussian random fields. We show that for scalar parameters optimal transport based SMC performs comparably to monomial based SMC but for Gaussian high-dimensional random fields optimal transport based SMC outperforms monomial based SMC. When comparing to ensemble Kalman inversion with mutation (EKI), we observe that for Gaussian random fields, optimal transport based SMC gives comparable or worse performance than EKI depending on the complexity of the parametrization. For non-Gaussian distributed parameters optimal transport based SMC outperforms EKI.

Keywords: parameter estimation, non-Gaussian posterior, tempering, particle approximation, Ensemble Transform Particle filter, Darcy flow

1. Introduction

We consider the inverse problem of inferring unknown parameters in models described by partial differential equations (PDEs), given incomplete noisy data/observations of the model outputs. We adopt the Bayesian approach where the unknowns are random functions with a prescribed prior measure that encompasses our prior statistical knowledge of the unknown. The solution to the Bayesian inversion problem is the posterior, i.e. the conditional distribution of the unknown parameters given the observed data. We can use the posterior to compute estimates of the unknown together with the degree of confidence in those estimates. We are interested in problems where the parameter-to-output map from the underlying PDE model is highly-nonlinear. These are particularly challenging problems since the resulting posterior cannot be obtained

analytically even when the prior and the noise distributions are assumed Gaussian. Hence, sampling methods are required to approximate (expectations under) the posterior which, in turn, is defined on a very high dimensional space after discretisation of the PDEs that define the forward problem.

Markov chain Monte Carlo (MCMC) is the method of choice to sample the Bayesian posterior [1]. In particular, there is a class of MCMC methods constructed in functional settings with mesh-invariant properties suitable for PDE-constrained identification problems [2]. However, the most standard version of these methods often exhibit excessively long correlations (e.g. up to 10^4 [3, 4]), a situation particularly exacerbated with highly-peaked (possibly multimodal) posteriors such as those arising when observational noise is small. Very long MCMC long chains (e.g. over 10^7 steps) are thus required to (i) ensure that MCMC fully explores the posterior measure thus capturing possibly multiple modes and (ii) produce sufficient independent samples to compute accurate posterior statistics. Since every step of MCMC involves at least one PDE solve, these methods become impractical for costly large-scale simulations. While more efficient MCMC can be used to approximate the posterior [5, 6], their proposals often required high-order derivatives of the likelihood which are not available in many applications where the simulator is accessible only in a black-box fashion.

Sequential Monte Carlo (SMC) samplers [7] offer a different sampling approach for approximating the Bayesian posterior. In the context of large-scale Bayesian inversion, adaptive SMC methods construct particle approximations of a sequence of intermediate measures that interpolate (e.g. via tempering) between the prior and the posterior. Particles and their weights are adapted on-the-fly to enable a controlled transition between those intermediate measures, thus facilitating to gradually move from a simple prior to a possibly complex posterior. The transition between two intermediate measures involves an importance resampling (IR) step by which the particles are weighted according to the tempered likelihood and then resampled according to those weights. This step is then followed by mutation of particles induced by sampling from a kernel with the IR measure as its invariant measure; this is typically conducted via running MCMC chains with the aforementioned target measure.

Adaptive SMC samplers for solving Bayesian inverse problems have been proposed in [4] and applied for the identification of the initial condition in the Navier-Stokes equations. This work showed that SMC can produce accurate approximations of the Bayesian posterior at a computational cost an order of magnitude smaller than those obtained via state-of-the-art MCMC. The same adaptive SMC sampler was used in [8] to infer permeability in a moving boundary problem arising in porous media flow. A theoretical framework for adaptive SMC framework was developed in [9] and tested numerically by inferring hydraulic conductivity in a groundwater flow model.

Despite of the computational advantages of using SMC samplers, their computational cost still poses severe limitations for its application to practical large-scale inverse problems. The cost of a single iteration (IR+mutation) within SMC is $J \times N_\mu$ where J is the number of particles and N_μ is the number of mutation MCMC moves.

Therefore, each iteration could involve over 10^4 PDE solves even for relatively small J and N_μ (i.e. $J = 10^3$ and $N_\mu = 10$). Hence, if the posterior is complex hence requiring several intermediate measures, the cost of SMC is prohibited unless high performance (HPC) resources are available to scale the cost of SMC with respect to J . While parallelisation is indeed one of the main advantages of SMC, the availability of HPC with $10^4 - 10^5$ processors for typical engineering and geophysical (practical) applications is the exception rather than norm. It is worth mentioning that reducing the cost of SMC via using small number of samples and/or reducing the number of mutation steps can be substantially detrimental to the accuracy of the particle approximation provided by SMC; see for example the work of [8] where SMC with limited number of particles ($10^2 - 10^3$) results in very poor approximations of the Bayesian posterior. Recent work aimed at reducing the computational cost of SMC samplers includes the development of multilevel versions [10, 11].

The aim of this work is to investigate the feasibility of an alternative, potentially more computationally affordable, approach to approximate the Bayesian posterior within the adaptive tempering SMC setting for Bayesian PDE-constrained inverse problems [4, 9]. The proposed approach consist of replacing the resampling step in SMC with a deterministic linear transformation that maps the system of particles that approximate two consecutive measures. At each iteration step within SMC, the transformation is obtained via solving an optimal transportation problem which, in turn, defines a deterministic coupling between two discrete random variables with realisations defined by the particles and with probabilities determined by their corresponding weights. The idea of replacing resampling by an optimal transformation within Bayesian algorithms was proposed in [12] where it was shown that the linear transport map leads to samples that converge to the posterior measures in large ensemble limit. In the context of data assimilation of partially observed dynamic systems, the idea of replacing IR by optimal transport maps is at the core of the so-called ensemble Transform Particle filter (ETPF) [12, 13]. The algorithm that we propose here for Bayesian inversion can then be partially viewed as the ETPF applied to artificial dynamics induced by the tempering scheme within SMC. In contrast to the approach of [12, 14], our work still relies on the mutation of particles according to a kernel that leaves each intermediate measure invariant.

Numerous work on data assimilation has shown that, when relatively small number of particles are used, ETPF provides more accurate state estimations compared to standard IR-based particles filters due to the sampling errors introduced by resampling. While methods such as ensemble Kalman filter (EnKF) can work well for small ensemble sizes compared to IR-based methods, they rely on Gaussian approximations which is often a severe limitation when the underlying distribution is, for example, multimodal. In contrast, the optimal transport within ETPF does not rely on Gaussian approximations and has been shown to be 1st order consistent for the mean, and to converge to the posterior measure in the large-ensemble size limit [12]. Here we investigate whether those well known advantages of ETPF can be exploited within the

setting of adaptive SMC for Bayesian inversion. As a proof-of-concept we apply the proposed algorithm to a Bayesian elliptic inverse problem arising in groundwater flow. The goal is to infer hydraulic conductivity from pressure measurements. We consider two parameterisations of the conductivity field aimed at assessing the method under two levels of complexity. In the first one we assume that the log-conductivity is a smooth function characterised by Gaussian random field under the prior. The second parameterisation consist of a channelised permeability that is described by a set of geometric parameters together with two random fields in the regions inside and outside the channel. While the first parameterisation yields posteriors which are relatively well approximated by Gaussians, the second parameterisation can result in multimodal distributions which are more difficult to capture with Gaussian approximations.

We compared the performance of the proposed technique against a fully resolved posterior computed by pcn-MCMC with sufficient steps to ensure that a chain is properly converged. We then compare the proposed technique against monomial based SMC as well as an ensemble Kalman inversion (EKI) technique that arises naturally from the adaptive SMC setting. This EKI methodology has been proposed in [15] as an alternative of [16]. Here this approach is modified to incorporate a mutation with the invariant measure.

2. Forward and Inverse Problem

Since we consider Bayesian inversion, it demands formulation of both a forward problem and an inverse problem. The forward problem consists of finding pressure from hydraulic conductivity. The "inverse" problem consists of two parts. First part is parametrization of hydraulic conductivity by a random variable. Second part is employment of the Bayes' rule to obtain the posterior distribution of the random variable from a given prior and a likelihood. The likelihood involves forward problem evaluation. Thus the Bayesian inversion employs the forward problem within the inverse problem.

2.1. Forward Model

The forward problem consist of the identification of the hydraulic conductivity, $\kappa(x)$, of a two-dimensional confined aquifer for which the physical domain is $D = [0, 6] \times [0, 6]$. Assuming that the flow within the aquifer is single-phase steady-state Darcy flow, the piezometric head $h(x)$, is given by the solution of [17]

$$-\nabla \cdot \kappa \nabla h = f \quad \text{in } D \quad (1)$$

where f represents recharge term. We use the Benchmark from [18, 19, 16] where f has the following form

$$f(x_1, x_2) = \begin{cases} 0 & \text{if } 0 < x_2 \leq 4, \\ 137 & \text{if } 4 < x_2 < 5, \\ 274 & \text{if } 5 \leq x_2 < 6. \end{cases} \quad (2)$$

and where the boundary conditions are given by

$$h(x_1, 0) = 100, \quad \frac{\partial h}{\partial x}(6, x_2) = 0, \quad -\kappa \frac{\partial h}{\partial x}(0, x_2) = 500, \quad \frac{\partial h}{\partial y}(x_1, 6) = 0. \quad (3)$$

We wish to infer $\kappa \in X := \{f \in L^\infty(D; \mathbb{R}) \mid \text{ess inf}_{x \in D} f(x) > 0\}$ from point observations of h collected at M locations denoted by $\{x_i\}_{i=1}^M \subseteq D$. To this end, we consider smoothed point observations defined by

$$\ell_j(h) = \int_D \frac{1}{2\pi\varepsilon^2} e^{-\frac{1}{2\varepsilon^2}(x-x_i)^2} h(x) dx$$

where $\varepsilon > 0$. Let us define the forward map $G : X \rightarrow \mathbb{R}^M$ by

$$G(\kappa) = (\ell_1(h), \dots, \ell_M(h)). \quad (4)$$

which maps permeability into predictions of hydraulic head at measurement locations. Assume that we have noisy measurements of $\{\ell_j(h)\}_{j=1}^M$ of the form

$$y_j = \ell_j(h) + \eta_j, \quad j = 1, \dots, M$$

where η_j represents measurement noise. Our aim is to reconstruct $\kappa \in X$ given $y = (y_1, \dots, y_M) \in \mathbb{R}^M$.

2.1.1. Parameterisation of permeability We consider the following two parameterisations of the permeability function $\kappa(x)$ that we wish to identify from observations of the Darcy flow model (1)-(3).

- P1: For the first model the parameter that we consider is simply the natural logarithm of κ , i.e. $u(x) = \log \kappa(x)$.
- P2: The second model consist of parameterisation of a piecewise continuous permeability of the form

$$\kappa(x) = \exp(u_1(x))\chi_{D_c}(x) + \exp(u_2(x))\chi_{D \setminus D_c}(x)$$

where $\kappa_1 = \exp(u_1(x))$ and $\kappa_2 = \exp(u_2(x))$ are continuous permeabilities inside and outside a sinusoidal channel with domain denoted by D_c . The geometry of the channel is parameterized by five parameters $\{d_i\}_{i=1}^5$ as described in Figure 1. The lower boundary of the channel is given by

$$x_2 = d_1 \sin(d_2 x_1 / 6) + \tan(d_3) x_1 + d_4$$

where we use the notation $x = (x_1, x_2) \in D$ in terms of the horizontal and vertical components. The upper boundary of the channel is given by $x_2 + d_5$. For this permeability model the parameters of interest are comprised in

$$u = (d_1, \dots, d_5, u_1, u_2)$$

where we assume that each d_i is restricted to an interval $A_i \equiv [d_i^-, d_i^+]$.

We define the following parameter space

$$U = \begin{cases} L^\infty(D; \mathbb{R}) & \text{for P1,} \\ \prod_{i=1}^5 A_i \times L^\infty(D; \mathbb{R}^2) & \text{for P2,} \end{cases}$$

with metric

$$|u|_U = \begin{cases} \|u\|_\infty & \text{for P1,} \\ \sum_{i=1}^5 |d_i| + \|u_1\|_\infty + \|u_2\|_\infty & \text{for P2,} \end{cases}$$

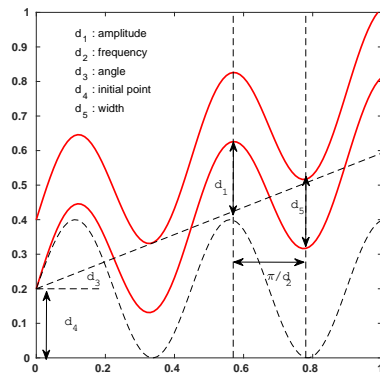


Figure 1. Geometrical configuration of channel flow.

The parameterizations described earlier define an abstract map $F : U \rightarrow X$ from the space of parameter to the space of admissible permeabilities, via

$$F(u) = \kappa. \quad (5)$$

We define the parameter-to-observations map $\mathcal{G} : U \rightarrow \mathbb{R}^M$ by $\mathcal{G} = G \circ F$ and reformulate the inverse problem (4) in terms of finding the parameter $u \in U$, given $y \in \mathbb{R}^M$ that satisfies

$$y = \mathcal{G}(u) + \eta \quad (6)$$

for $\eta = (\eta_1, \dots, \eta_M) \in \mathbb{R}^M$. The continuity of the parameter-to-observations map \mathcal{G} for this, and more general cases, has been established in [20, 3].

2.2. The Bayesian Inverse Problem

In order to address the inverse problem formulated via (6) we adopt the Bayesian framework [20] where η is a random vector and u is a random function. We put a prior, $\mu_0(u)$, on the unknown u , and define the random variable $y|u$ under the standard assumption that $\eta \sim N(0, \sigma^2 I)$ independent of u . The solution to the inverse problem in the Bayesian setting is the posterior measure on $u|y$. In the following sections we introduce the prior and likelihood which by the infinite-dimensional framework of [20] ensure that the posterior measure exists and is continuous with respect to appropriate metrics.

2.2.1. The Prior For P1 we consider Gaussian prior $\mu_0 = N(m, C)$ with mean m and covariance C . We define C via a correlation function given by the Wittle-Matern correlation function defined by [21]:

$$c(x, y) = \sigma_0^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{|x - y|}{l} \right)^\nu K_\nu \left(\frac{|x - y|}{l} \right), \quad (7)$$

where Γ is the gamma function, l is the characteristic length scale, σ_0^2 is an amplitude scale and K_ν is the modified Bessel function of the second kind of order ν . The parameter ν controls the regularity of the samples.

For P2 we assume independence between geometric parameters and log-permeabilities and thus consider a prior of the form

$$\mu_0(du) = \prod_{i=1}^5 \pi_0^{A_i}(d_i) \otimes N(m_1, C_1) N(m_2, C_2) \quad (8)$$

where $\pi_0^A(x)$ is the uniform density defined by

$$\pi_0^A(x) = \begin{cases} \frac{1}{|A|} & x \in A, \\ 0 & x \notin A. \end{cases} \quad (9)$$

In expression (8) $N(m_1, C_1)$ and $N(m_2, C_2)$ are two Gaussians such as those described earlier in terms of the correlations function from (7).

2.2.2. The likelihood We assume the unknown u is independent of the observational noise $\eta \sim N(0, \sigma^2)$. We note that $y|u \sim N(\mathcal{G}(u), \sigma^2 I)$, hence the likelihood is given by

$$l(u, y) \propto \exp(-\Phi(u, y)) \quad (10)$$

where $\Phi(u, y)$ is the data misfit defined by

$$\Phi(u, y) = \frac{1}{2\sigma^2} \|y - \mathcal{G}(u)\|^2 \quad (11)$$

2.2.3. The Posterior The selection of prior measures from subsection 2.2.1 satisfies that $\mu_0(U) = 1$; i.e. samples from μ_0 are in U almost surely [20, 3]. This property, together with the continuity of the forward map defined in subsection 2.1, can be used in the Bayesian framework of [20, 3] to conclude that (i) the posterior measure $\mu(u)$ on $u|y$ exists and is absolutely continuous with respect to the prior; and(ii) μ_0 and has a density with respect to μ_0 given by the following Bayes' rule

$$\frac{d\mu}{d\mu_0} = \frac{1}{Z} l(u, y) \quad (12)$$

where

$$Z = \int_U l(u, y) \mu_0(du) \quad (13)$$

3. Sequential Monte Carlo for Bayesian inversion

Since we consider a highly nonlinear model, an iterative approach to Bayesian inversion is essential. In the framework of SMC it is performed by tempering (or annealing), when the prior measure bridged to the posterior measure not at once but through tempered measures. It should be noted that the number of tempered measures is not predefined, which could be a potential computational burden. In order to avoid filter degeneracy both resampling and mutation (or jittering) has to be performed. In the "classical" approach we perform monomial resampling, which we propose to replace by resampling based on optimal transport.

3.1. Adaptive SMC

The SMC approach to Bayesian inversion involves bridging the prior μ_0 and the posterior μ via a sequence of intermediate artificial measures $\{\mu_n\}_{n=0}^N$, with $\mu_N = \mu$, defined by

$$\frac{d\mu_n}{d\mu_0}(u) \propto l_n(u, y) \equiv l(u, y)^{\phi_n} \quad (14)$$

where $\{\phi_n\}_{n=0}^N$ is a set of tempering parameters that satisfy $0 = \phi_0 < \phi_1 < \dots < \phi_N = 1$. Expression (14) formally implies

$$\frac{d\mu_n}{d\mu_{n-1}}(u) = \frac{1}{Z_n} l(u, y)^{(\phi_n - \phi_{n-1})} \quad (15)$$

where

$$Z_n \equiv \int_X l(u, y)^{(\phi_n - \phi_{n-1})} \mu_{n-1}(du) \quad (16)$$

Let us then assume that at the iteration level $n - 1$, the tempering parameter ϕ_{n-1} has been specified, and that a set of particles $\{u_{n-1}^{(j)}\}_{j=1}^J$ provides the following approximation (with equal weights) of the intermediate measure μ_{n-1} :

$$\mu_{n-1}^J(u) \equiv \frac{1}{J} \sum_{j=1}^J \delta_{u_{n-1}^{(j)}}(u) \simeq \mu_{n-1}(u). \quad (17)$$

Then from (16) it follows that

$$Z_n \simeq \sum_{j=1}^J l(u_{n-1}^{(j)}, y)^{(\phi_n - \phi_{n-1})} \quad (18)$$

and thus, for any measurable f , we have that

$$\begin{aligned} \mathbb{E}^{\mu_n}(f(u)) &\equiv \int_X f(u) \mu_n(du) = \frac{1}{Z_n} \int_X f(u) l(u, y)^{(\phi_n - \phi_{n-1})} \mu_{n-1}(du) \\ &\simeq \left[\sum_{j=1}^J l(u_{n-1}^{(j)}, y)^{(\phi_n - \phi_{n-1})} \right]^{-1} \sum_{j=1}^J l(u_{n-1}^{(j)}, y)^{(\phi_n - \phi_{n-1})} f(u_{n-1}^{(j)}), \\ &= \sum_{j=1}^J W_n^{(j)} f(u^{(j)}), \end{aligned} \quad (19)$$

where the importance weights for the approximation of μ_n are given by

$$W_n^{(j)} = \mathcal{W}_{n-1}^{(j)}[\phi_n] \equiv \frac{l(u_{n-1}^{(j)}, y)^{\phi_n - \phi_{n-1}}}{\sum_{s=1}^J l(u_{n-1}^{(s)}, y)^{\phi_n - \phi_{n-1}}}. \quad (20)$$

From (19) we see that the importance (normalized) weights $W_n^{(j)}$ assigned to each particle $u_{n-1}^{(j)}$ define the following empirical (particle) approximation of μ_n :

$$\mu_n^J(u) \equiv \sum_{j=1}^J W_n^{(j)} \delta_{u_{n-1}^{(j)}}(u). \quad (21)$$

3.1.1. Selection-Resampling Step From the previous subsection it follows that adaptive SMC requires then to select the tempering parameters ϕ_n so that the two consecutive measures μ_{n-1} and μ_n are sufficiently closed for the IS approximating to be accurate. To this end, a common procedure [22] involves imposing a threshold on the effective sample size (ESS) defined by

$$\text{ESS}_n(\phi) \equiv \left[\sum_{j=1}^J (\mathcal{W}_{n-1}^{(j)}[\phi])^2 \right]^{-1} \quad (22)$$

which, in turn, provides a measure of the quality of the population. In other words, ϕ_n is defined by the solution to

$$\text{ESS}_n(\phi) = J_{\text{thresh}}, \quad (23)$$

for a user-defined parameter J_{thresh} on the ESS. A bisection algorithm on the interval $(\phi_{n-1}, 1]$ can be used to solve (23) [16]. If $\text{ESS}_n(1) > J_{\text{thresh}}$, then then we can simply set $\phi_n = 1$ as no further tempering is thus required.

Once the tempering parameter ϕ_n has been computed via (23), normalised weights (20) can be computed. Since some of these can be very low, resampling with replacement according to these weights is then required to discard particles associated with those low weights. After resampling, a new set of equally-weighted particles denoted by $\hat{u}_n^{(j)}$ ($j = 1, \dots, J$) provide a particle approximation of the measure μ_n .

3.1.2. Mutation Phase In order to add diversity to the resampled particles $\hat{u}_n^{(j)}$ computed in the selection-resampling step, a mutation step is included in most SMC methodologies. This mutation consists of sampling from a Markov kernel \mathcal{K}_n with invariant distribution μ_n . This can be achieved by running N_μ steps of an MCMC algorithm that has target distribution equal to μ_n . An example of MCMC suitable for the parameterisation P1 of section 2.1.1 is the preconditioned Crank-Nicolson (pcn)-MCMC [2] displayed in Algorithm 1. This algorithm samples from the target μ_n with reference measure $\mu_0 = N(m, C)$; we recall these two measures are related by (12). The resulting particles denoted by $\{u_n^{(j)}\}_{j=1}^J$ ($u_n^{(j)} \sim \mathcal{K}_n(\hat{u}_n^{(j)}, \cdot)$) provide a particle approximation of μ_n in the form

$$\mu_n^J \equiv \frac{1}{J} \sum_{j=1}^J \delta_{u_n^{(j)}} \quad (24)$$

Convergence of (24) to μ_n in the large ensemble size limit can be found in [9]. The complete adaptive SMC sampler is displayed in Algorithm 2.

Algorithm 1 pcn-MCMC to generate samples from a μ_n -invariant Markov kernel with $\mu_0 = N(m, C)$

Select $\beta \in (0, 1)$ and an integer N_μ .

for $j = 1, \dots, J$ **do**

Initialize $\nu^{(j)}(0) = \hat{u}_n^{(j)}$

while $\alpha \leq N_\mu$ **do**

(1) **pcN proposal.** Propose u_{prop} from

$$u_{\text{prop}} = \sqrt{1 - \beta^2} \nu^{(j)}(\alpha) + (1 - \sqrt{1 - \beta^2})m + \beta\xi, \quad \text{with } \xi \sim N(0, C)$$

(2) Set $\nu^{(j)}(\alpha+1) = u_{\text{prop}}$ with probability $a(\nu^{(j)}(\alpha), u)$ and $\nu^{(j)}(\alpha+1) = \nu^{(j)}(\alpha)$ with probability $1 - a(\nu^{(j)}(\alpha), u)$, where

$$a(u, v) = \min \left\{ 1, \frac{l(u, y)^{\phi_n}}{l(v, y)^{\phi_n}} \right\}$$

with l defined in (10)

(3) $\alpha \leftarrow \alpha + 1$

end while

end for

Algorithm 2 SMC algorithm for High-Dimensional Inverse Problems

Let $\{u_0^{(j)}\}_{j=1}^J \sim \mu_0$ be the initial ensemble of J particles.

Define the tunable parameters J_{thresh} and N_μ .

Set $n = 0$ and $\phi_0 = 0$

while $\phi_n < 1$ **do**

$n \rightarrow n + 1$

Compute the likelihood (10) $l(u_{n-1}^{(j)}, y)$ (for $j = 1, \dots, J$)

Compute the tempering parameter ϕ_n :

if $\min_{\phi \in (\phi_{n-1}, 1)} \text{ESS}_n(\phi) > J_{\text{thresh}}$ **then**

set $\phi_n = 1$.

else

compute ϕ_n such that $\text{ESS}_n(\phi) \approx J_{\text{thresh}}$

using a bisection algorithm on $(\phi_{n-1}, 1]$.

end if

Computing weights from expression (20) $W_n^{(j)} \equiv \mathcal{W}_{n-1}^{(j)}[\phi_n]$

Resample. Let $(p^{(1)}, \dots, p^{(N_p)}) \in \mathcal{R}(W_n^{(1)}, \dots, W_n^{(N_p)})$,

Set $\hat{u}_n^{(j)} \equiv u_{n-1}^{(p^{(j)})}$ and $W_n^{(j)} = \frac{1}{J}$

Mutation. Sample $u_n^{(j)} \sim \mathcal{K}_n(\hat{u}_n^{(j)}, \cdot)$ via Algorithm 1.

end while

Approximate μ_n by $\mu_n^J \equiv \frac{1}{J} \sum_{j=1}^J \delta_{u_n^{(j)}}$

3.2. Optimal Transport within SMC

In this section we assume that $X = \mathbb{R}^K$. We denote U_{n-1} a discrete random variable with realisations $\{u_{n-1}^{(j)}\}_{j=1}^J$ and probabilities $\{W_n^{(j)}\}_{j=1}^J$. We denote U_n the random variable with samples $\{\hat{u}_{n-1}^{(j)}\}_{j=1}^J$ with equal weights. The aim is to replace the resampling step in the method above with resampling that maximizes the covariance between U_{n-1} and U_n . Such a resampling is performed by finding a coupling between the posterior defined by the weights $\{W_n^{(j)}\}_{j=1}^J$ and the uniform probability density such that it maximizes the covariance between U_{n-1} and U_n . It appeared that such a coupling simultaneously minimizes the expectation between $\|u_{n-1} - \hat{u}_n\|^2$ and is defined as the solution of the Monge-Kantorovitch problem with cost function $c(u_{n-1}, \hat{u}_n) = \|u_{n-1} - \hat{u}_n\|^2$.

Thus the above described coupling is a $J \times J$ matrix T^* with non-negative entries T_{ij}^* that satisfy

$$\sum_{i=1}^J T_{ij}^* = \frac{1}{J}, \quad \sum_{j=1}^J T_{ij}^* = W_i, \quad (25)$$

and minimizes

$$\sum_{i,j=1}^J T_{ij} \|u_{n-1}^{(i)} - \hat{u}_n^{(j)}\|^2 \equiv \sum_{i,j=1}^J T_{ij} \|u_{n-1}^{(i)} - u_{n-1}^{(j)}\|^2 \quad (26)$$

for T_{ij}^* . This is a *linear* transport problem of finding J^2 unknowns. Then the linear transformation gives new samples according to

$$\hat{u}_n^{(j)} := \sum_{i=1}^J P_{ij} u_{n-1}^{(j)} \quad \text{for } j = 1, \dots, J, \quad (27)$$

where $P_{ij} = JT_{ij}^*$.

The deterministic optimal transformation (27) converges weakly to the solution of the underlying continuous Monge-Kantorovitch problem as $J \rightarrow \infty$ [12]. ETPF is first order consistent, since

$$\bar{\hat{u}}_n = \frac{1}{J} \sum_{j=1}^J \hat{u}_n^{(j)} = \frac{1}{J} \sum_{j=1}^J \sum_{i=1}^J P_{ij} u_{n-1}^{(j)} = \sum_{j=1}^J \sum_{i=1}^J T_{ij}^* u_{n-1}^{(j)} = \sum_{j=1}^J W_n^{(j)} u_{n-1}^{(j)}.$$

There also exists a second-order accurate ETPF [23], which however does not satisfy $T_{ij}^* \geq 0$. The main difference between resampling based on optimal transport and monomial resampling is that the former one is optimal in the sense of the Monge-Kantorovitch problem, while the latter one is non-optimal in that sense.

The computational complexity of finding the minimizer of (27) is in general $O(J^3 \ln J)$, which has been reduced to $O(J^2 \ln J)$ in [24]. It can be further improved by employing fast iterative methods for finding approximate minimizers using the Sinkhorn distance [25], which was implemented in [23] for the second-order accurate ETPF. The algorithm of Earth's moving distances of [24] is available as both *MATLAB* and *Python* codes and is used here. The complete adaptive optimal transport based SMC sampler is displayed in Algorithm 3.

Algorithm 3 Optimal transport based SMC algorithm for High-Dimensional Inverse Problems

Let $\{u_0^{(j)}\}_{j=1}^J \sim \mu_0$ be the initial ensemble of J particles.
 Define the tunable parameters J_{thresh} and N_μ .
 Set $n = 0$ and $\phi_0 = 0$
while $\phi_n < 1$ **do**
 $n \rightarrow n + 1$
 Compute the likelihood (10) $l(u_{n-1}^{(j)}, y)$ (for $j = 1, \dots, J$)
 Compute the tempering parameter ϕ_n :
 if $\min_{\phi \in (\phi_{n-1}, 1)} \text{ESS}_n(\phi) > J_{\text{thresh}}$ **then**
 set $\phi_n = 1$.
 else
 compute ϕ_n such that $\text{ESS}_n(\phi) \approx J_{\text{thresh}}$
 using a bisection algorithm on $(\phi_{n-1}, 1]$.
 end if
 Computing weights from expression (20) $W_n^{(j)} \equiv \mathcal{W}_{n-1}^{(j)}[\phi_n]$
 Resample based on optimal transport. Compute $d_{ij} = \|u_{n-1}^{(i)} - u_{n-1}^{(j)}\|^2$ (for $i, j = 1, \dots, J$). Supply $\{d_{ij}\}_{i,j=1}^J$ and $\{W_n^{(j)}\}_{j=1}^J$ to the Earth's moving distances algorithm of Pele & Werman. The output is the coupling $\{T_{ij}^*\}_{i,j=1}^J$.
 Compute new samples $\hat{u}_n^{(j)}$ (27) and set $W_n^{(j)} = \frac{1}{J}$.
 Mutation. Sample $u_n^{(j)} \sim \mathcal{K}_n(\hat{u}_n^{(j)}, \cdot)$ via Algorithm 1.
end while
 Approximate μ_n by $\mu_n^J \equiv \frac{1}{J} \sum_{j=1}^J \delta_{u_n^{(j)}}$

3.3. Gaussian Approximation of SMC via ensemble Kalman inversion

A natural approximation that arises from the adaptive SMC framework described in subsection 3.1 involves ensemble Kalman inversion (EKI) [8]. More specifically, let us assume that at the $n - 1$ iteration level, we approximate μ_{n-1} with a Gaussian $\hat{\mu}_{n-1} = N(m_{n-1}, C_{n-1})$ where the mean m_{n-1} and covariance C_{n-1} are the empirical mean and covariance of the particles (assumed with equal weights) at the current iteration level. That is,

$$m_{n-1} \equiv \frac{1}{J} \sum_{j=1}^J u_{n-1}^{(j)}, \quad C_{n-1} \equiv \frac{1}{J-1} \sum_{j=1}^J (u_{n-1}^{(j)} - m_{n-1}) \otimes (u_{n-1}^{(j)} - m_{n-1}) \quad (28)$$

If we now linearise the forward map around m_{n-1} and replace Frechet derivatives of the forward map with covariances/crosscovariances as in [16], it can be shown that the application to Bayes rule yields an approximate posterior $\hat{\mu}_n = N(m_n, C_n)$ with mean and covariance given by

$$m_n = m_{n-1} + C_{n-1}^{u\mathcal{G}} (C_{n-1}^{\mathcal{G}\mathcal{G}} + \alpha_n \Gamma)^{-1} (y - \bar{\mathcal{G}}_{n-1}), \quad (29)$$

$$C_n = C_{n-1} - C_{n-1}^{u\mathcal{G}} (C_{n-1}^{\mathcal{G}\mathcal{G}} + \alpha_n \Gamma)^{-1} C_{n-1}^{\mathcal{G}u}, \quad (30)$$

where

$$\bar{\mathcal{G}}_{n-1} \equiv \frac{1}{J} \sum_{j=1}^J \mathcal{G}(u_{n-1}^{(j)}), \quad C_{n-1}^{u\mathcal{G}} \equiv \frac{1}{J-1} \sum_{j=1}^J (u_{n-1}^{(j)} - m_{n-1}) \otimes (\mathcal{G}(u_{n-1}^{(j)}) - \bar{\mathcal{G}}_{n-1}), \quad (31)$$

$$C_{n-1}^{\mathcal{G}\mathcal{G}} \equiv \frac{1}{J-1} \sum_{j=1}^J (\mathcal{G}(u_{n-1}^{(j)}) - \bar{\mathcal{G}}_{n-1}) \otimes (\mathcal{G}(u_{n-1}^{(j)}) - \bar{\mathcal{G}}_{n-1}), \quad (32)$$

and where

$$\alpha_n = \frac{1}{\phi_n - \phi_{n-1}}. \quad (33)$$

Since we are interested in a particle approximation of $\hat{\mu}_n = N(m_n, C_n)$, we can use the following expression

$$\hat{u}_n^{(j)} = u_{n-1}^{(j)} + C_{n-1}^{u\mathcal{G}} (C_{n-1}^{\mathcal{G}\mathcal{G}} + \alpha_n \Gamma)^{-1} (y_n^{(j)} - \mathcal{G}_n(u_{n-1}^{(j)})), \quad (34)$$

where

$$y_n^{(j)} \equiv y + \eta_n^{(j)}, \quad \eta_n^{(j)} \sim N(0, \alpha_n \Gamma_n). \quad (35)$$

Standard Kalman filter arguments [26] can be used to show the particle approximation provided by (34)-(35) converges to $\hat{\mu}_n$ as $J \rightarrow \infty$. We note in passing that, within the adaptive SMC framework used here, the regularisation/inflation parameter α_n in formulas (33) is computed based on the ESS criteria discussed in subsection 3.1.1.

It is important to emphasize that, in general, the approximate Gaussian measure $\hat{\mu}_n$ coincides with μ_n only when the forward map is linear and the prior μ_0 is Gaussian. The approximation provided by EKI will deteriorate when we depart from Gaussian-linear assumptions. Therefore, we propose to conduct MCMC mutations to each of the particles in (34) with the aim of improving the approximation of each posterior measure μ_n . The complete EKI-based algorithm is displayed in Algorithm 4. We recognise that this is only an ad-hoc approach which can be potentially improved via more rigorous versions approaches. For example, there exists data assimilation approaches that use ensemble Kalman filter (EnKF) as proposal in the importance sampling step within SMC [27].

Algorithm 4 EKI approximation to SMC

Let $\{u_0^{(j)}\}_{j=1}^J \sim \mu_0$ be the initial ensemble of J particles.
 Define the tunable parameters J_{thresh} and N_μ .
 Set $n = 0$ and $\phi_0 = 0$
while $\phi_n < 1$ **do**
 $n \rightarrow n + 1$
 Compute the likelihood (10) $l(u_{n-1}^{(j)}, y)$ (for $j = 1, \dots, J$)
 Compute the tempering parameter ϕ_n :
 if $\min_{\phi \in (\phi_{n-1}, 1)} \text{ESS}_n(\phi) > J_{\text{thresh}}$ **then**
 set $\phi_n = 1$.
 else
 compute ϕ_n such that $\text{ESS}_n(\phi) \approx J_{\text{thresh}}$
 using a bisection algorithm on $(\phi_{n-1}, 1]$.
 end if
 Generate particles $\{\hat{u}_n^{(j)}\}_{j=1}^J$ according to (34).
 Mutation. Sample $u_n^{(j)} \sim \mathcal{K}_n(\hat{u}_n^{(j)}, \cdot)$ via Algorithm 1.
end while
 Approximate μ_n by $\mu_n^J \equiv \frac{1}{J} \sum_{j=1}^J \delta_{u_{n,r}^{(j)}}$

4. Numerical experiments

In this section we perform numerical experiments to infer P1 and P2 parameters. We compare optimal transport based SMC to both monomial based SMC and EKI, which we denote optimal, monomial, and Kalman, respectively. We analyze methods performance with respect to a pcn-MCMC solution, which we denote as reference. We combine 50 independent chains each of the length 10^6 and 10^5 burn-in period and thinning 10^3 .

Observations of pressure were obtained from the true permeability with observation noise from normal distribution with zero mean and standard deviation of 2% of the true pressure. We should note that both the true random variable and an initial ensemble of parameterized permeability are drawn from the same prior distribution as the prior includes knowledge about geological properties. However, the true solution is computed on a fine grid and an initial guess on a coarse grid, which is half the resolution of the fine grid. The uncertain parameter for P1 inference has the dimension of the coarse grid, i.e. $4900 = 70^2$. The uncertain parameter for P2 inference has the dimension of the coarse grid twice, due to permeability defined inside and outside channel but on the whole grid, plus the dimension of the geometrical parameters, i.e. $5005 = 50^2 + 50^2 + 5$.

For log-permeability parameters, the prior is normal distribution with mean 5 for P1, and for P2 with mean 15 outside channel and 100 inside channel. For geometrical parameters, the prior is uniform: $d_1 \sim U[0.05 \times 6, 0.35 \times 6]$, $d_2 \sim U[\pi/2, 6\pi]$, $d_3 \sim U[-\pi/2, \pi/2]$, $d_4 \sim U[0, 6]$, $d_5 \sim U[0.02 \times 6, 0.7 \times 6]$. For tempering we choose the effective ensemble size threshold $J_{\text{thresh}} = J/3$ and for mutations the length of Markov

chain $N_\mu = 10$ to save computational costs. For P2, we use Metropolis-within-Gibbs methodology of [3] to separate geometrical parameters and log-permeability parameters within the mutation step, since it allows to better exploit the structure of the prior. The proposal design for the geometric parameters within the Metropolis-within-Gibbs consist of local moves within the intervals of the prior with a step size that we tune to achieve acceptance rates between 20% and 30%. Geometrical parameters that fall outside those intervals are projected back via a projection that preserves reversibility of the proposal with respect to the prior [3]. We perform numerical experiments with different ensemble sizes of 100, 500, and 1000. We perform 10 simulations with different realizations of the initial ensemble to check the robustness of results.

For log-permeability, we compute L^2 norm of the error in the mean with respect to the reference

$$\text{Error} = \|\bar{u} - \bar{u}^{\text{ref}}\|, \quad \text{where} \quad \bar{u} = \frac{1}{J} \sum_{j=1}^J u^{(j)}.$$

For the geometrical parameters of P2 parametrization, we compute Kullback-Leibler divergence with respect to the reference

$$D_{\text{KL}}(p^{\text{ref}} \parallel p) = \sum_{i=1}^{J_b} p^{\text{ref}}(u_i) \log \frac{p^{\text{ref}}(u_i)}{p(u_i)}, \quad (36)$$

where $J_b = J/10$ is chosen number of bins and $p(u_i)$ is approximated by the weights.

4.1. Numerical inference for P1

For P1, we perform a numerical experiment using 36 uniformly distributed observations. In Figure 2, we plot error in the mean log-permeability with respect to reference. We observe that while optimal transport based SMC outperforms monomial based SMC for all ensemble sizes, EKI outperforms both SMC methods. This is due to the nature of P1 parametrization and only two degrees of freedom (mean and variance) of EKI. In Figure 3, we plot mean log-permeability for a simulation with smallest error at ensemble size 100 and reference mean log-permeability. We see that monomial based SMC gives a less smooth estimation compared to optimal transport based SMC, EKI, and reference, which leads to larger error.

For ensemble sizes considered here, the number of tempering steps on average is 15 for optimal transport based SMC, and 17 for both monomial based SMC and EKI. Thus in terms of computational cost optimal transport based SMC is equivalent to monomial based SMC, since computational complexity of the forward model is higher than $O(J \ln J)$.

4.2. Numerical inference for P2

For P2, we perform a numerical experiment using 9 uniformly distributed observations. For ensemble size considered here, the number of tempering steps on average is 8 for EKI,

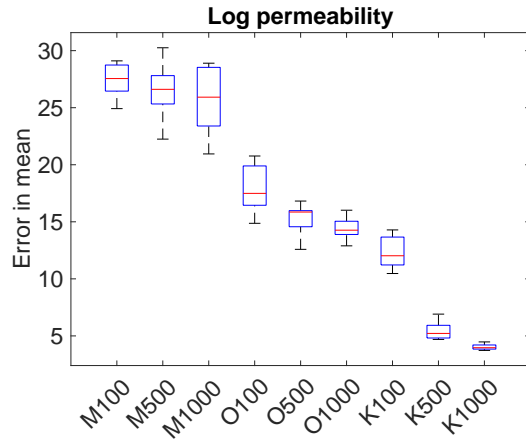


Figure 2. Box plot of the error in the mean log-permeability for P1 inference. Central mark is the median, edges of the box are the 25th and 75th percentiles, whiskers extend to the most extreme datapoints. On x-axis numbers stand for ensemble sizes, M stands for monomial based SMC, O for optimal transport based SMC, and K for EKI.

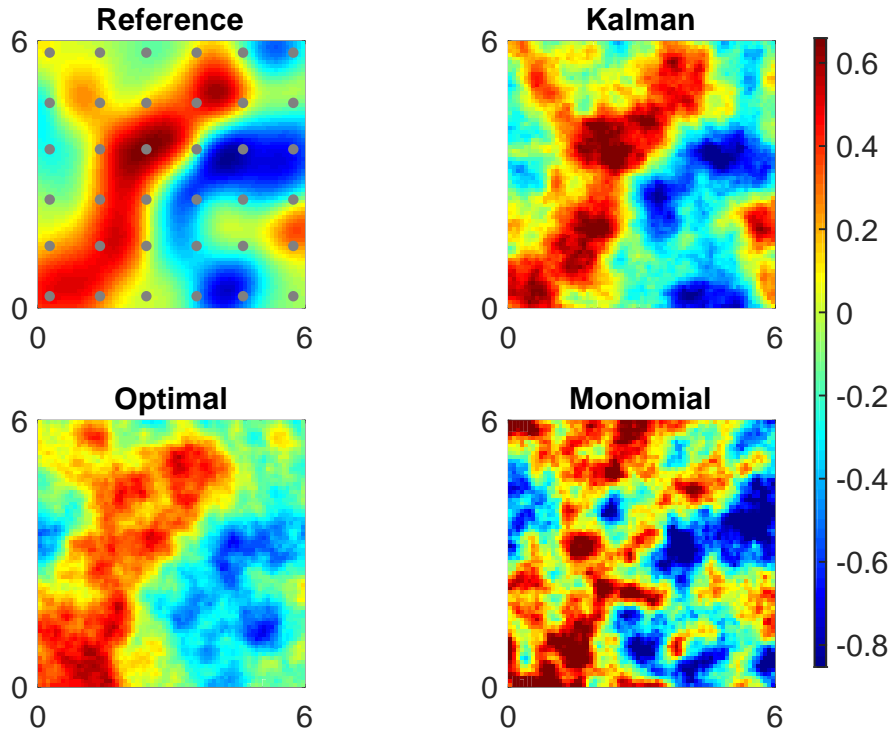


Figure 3. Mean log-permeability for P1 inference for the lowest error at ensemble size 100. Observation locations are shown in circles.

and 7 for both optimal transport based SMC and monomial based SMC. In Figure 4, we plot error in the mean log-permeability with respect to reference for permeability outside channel on the left and for permeability inside channel on the right. We observe that while optimal transport based SMC still outperforms monomial based SMC for

all ensemble sizes, it is now comparable to EKI. This is due to a small number of observations. In Figures 5–6, we plot mean log-permeability for a simulation with

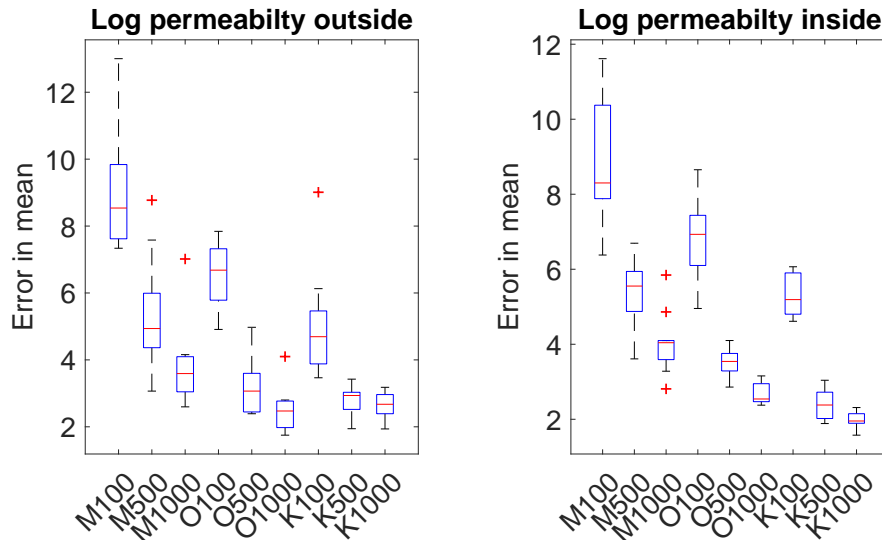


Figure 4. Box plot of the error in the mean log-permeability for P2 inference. Central mark is the median, edges of the box are the 25th and 75th percentiles, whiskers extend to the most extreme datapoints, and crosses are outliers. On the left: outside channel, on the right: inside channel. On x-axis numbers stand for ensemble sizes, M stands for monomial based SMC, O for optimal transport based SMC, and K for EKI.

smallest error at ensemble size 100 and reference mean log-permeability for permeability outside channel and for permeability inside channel, respectively. We see that monomial based SMC gives a less smooth estimation compared to optimal transport based SMC, EKI, and reference, which leads to larger error.

In Figure 7, we show posterior estimations of geometrical parameters. We see that all the parameters except amplitude and width exhibit strongly non-Gaussian behaviour. In Figure 8, we show a trace plot of frequency from a chain of the reference to check whether two modes are being sampled within each chain. We observe that the chain is properly mixed.

In Figure 9, we plot KL divergence for geometrical parameters. We observe that EKI performs better than optimal transport based SMC for amplitude and width, while worse for other parameters. Monomial based SMC performs comparably to optimal transport based SMC though not consistently better or worse. We should recall, however, that optimal transport based SMC outperforms monomial based SMC for log-permeability both inside and outside channel. In Figure 10, we show mean field of permeability over the channelized domain for the lowest error at ensemble size 1000.

5. Conclusions

Accurate estimation of the posterior distribution of uncertain model parameters of strongly nonlinear problems remains a challenging problem. Parameters are high

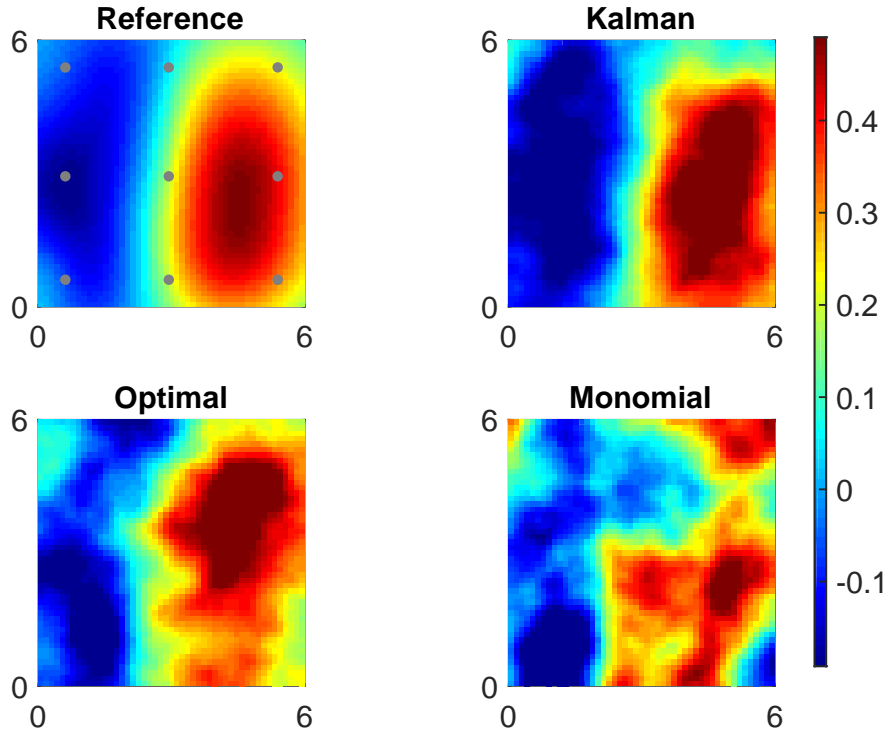


Figure 5. Mean log-permeability outside channel for P2 inference for the lowest error at ensemble size 100. Observation locations are shown in circles.

dimensional, they are not observed, and they do not have a dynamical equation. Moreover, due to nonlinearity of models even Gaussian prior of parameters might result in non-Gaussian posterior. Since MCMC is computationally unfeasible for high-dimensional problems, adaptive SMC is an alternative to estimate posterior distributions in the Bayesian framework. However, adaptive SMC still requires large ensembles.

In order to decrease ensemble size, we proposed to introduce optimal transport based resampling from [12] to adaptive SMC. Optimal transport based resampling creates new samples by maximizing variance between prior and posterior. It has been already shown for state estimation and parameter estimation with low dimension, that particle filter with optimal transport based resampling outperforms particle filter with monomial based resampling. As it was aimed to estimate time-evolving model states of chaotic systems, simple inflation was sufficient to mutate particles.

Here we have adopted optimal transportation to elliptic Bayesian inverse problems. We have shown that optimal transport based SMC has a high potential for Bayesian inversion of high-dimensional parameters. It outperforms EKI for non-Gaussian distributed scalar parameters, and monomial based SMC for high-dimensional parameters. The latter becomes even more pronounced with more measurements. However, for scalar parameters optimal transport based SMC is not consistently better than monomial based SMC, thus requiring more detailed sensitivity analysis.

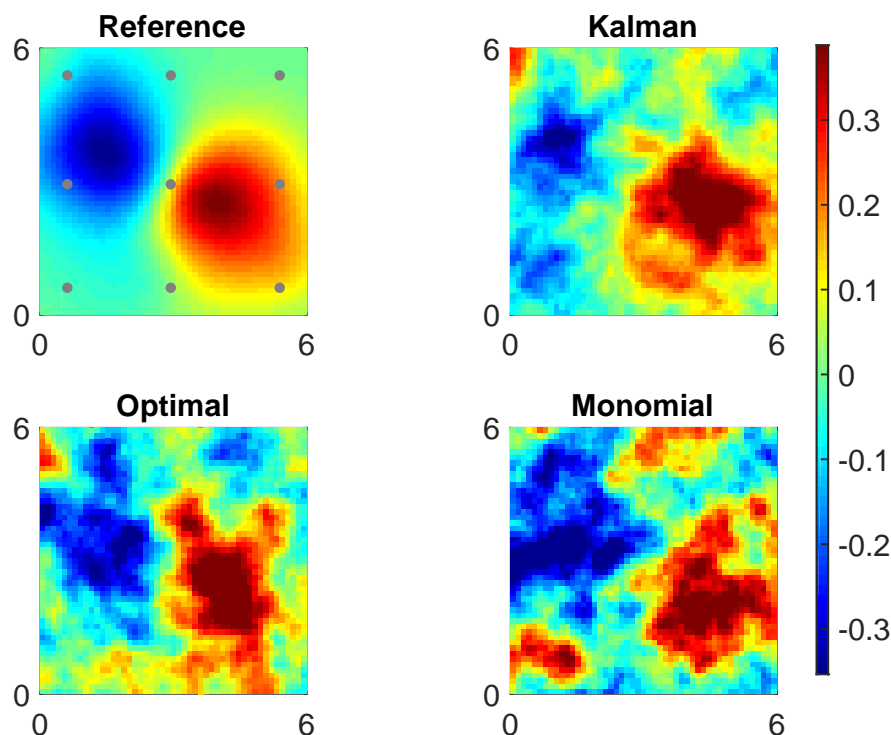


Figure 6. Mean log-permeability inside channel for P2 inference for the lowest error at ensemble size 100. Observation locations are shown in circles.

Moreover, optimal transport based SMC still underestimates variance (not shown), which could be improved by considering second order consistent optimal transport resampling instead of first order. However, second order consistent optimal transport resampling does not necessary provide with non-negative transformations. Finally, optimal transport resampling does not need to be restricted to finite dimensions, at least theoretically [28], with the challenge of finding such a minimizer computationally.

Acknowledgments

This work is part of the research programme Shell-NWO/FOM Computational Sciences for Energy Research (CSER) with project number 14CSER007 which is partly financed by the Netherlands Organization for Scientific Research (NWO).

- [1] Kaipio J and Somersalo E 2005 *Statistical and computational inverse problems* (Springer Science+Business Media, Inc.)
- [2] Cotter S, Roberts G, Stuart A and White D 2013 *Statistical Science* **28** 424–446
- [3] Iglesias M, Lin K and Stuart A 2014 *Inverse Problems* **30** 114001
- [4] Kantas N, Beskos A and Jasra A 2014 *SIAM/ASA Journal Uncertainty Quantification* **2** 464–489
- [5] Bui-Thanh T and Girolami M 2014 *Inverse Problems* **30** 114014
- [6] Martin J, Wilcox L, Burstedde C and Ghattas O 2012 *SIAM Journal on Scientific Computing* **34** A1460–A1487

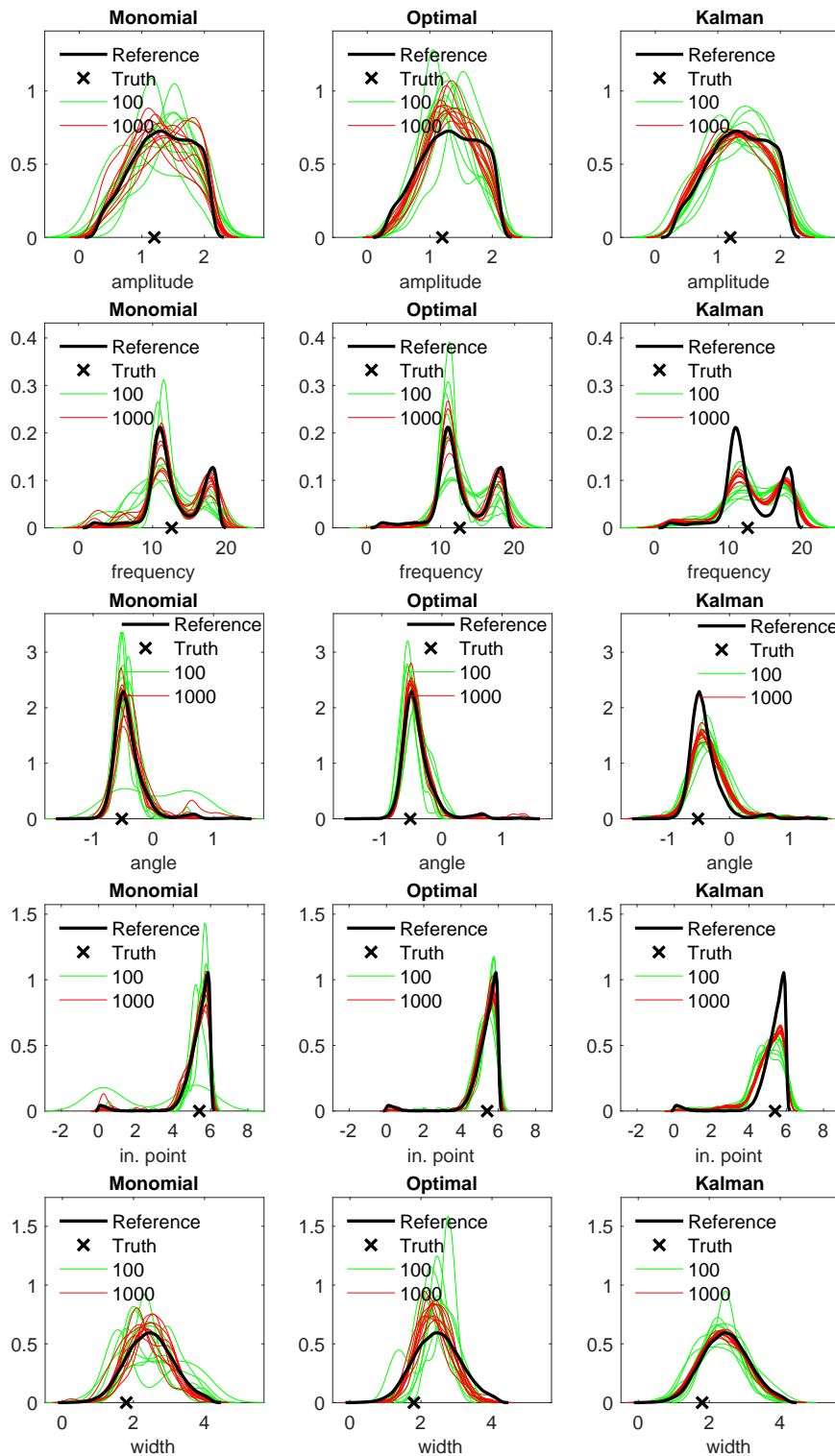


Figure 7. Posterior of geometrical parameters for P2 inference. In black is reference, in green 10 simulations of ensemble size 100, in red 10 simulations of ensemble size 1000. The true parameters are shown as black cross.

[7] Del Moral P, Doucet A and Jasra A 2006 *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68** 411–436

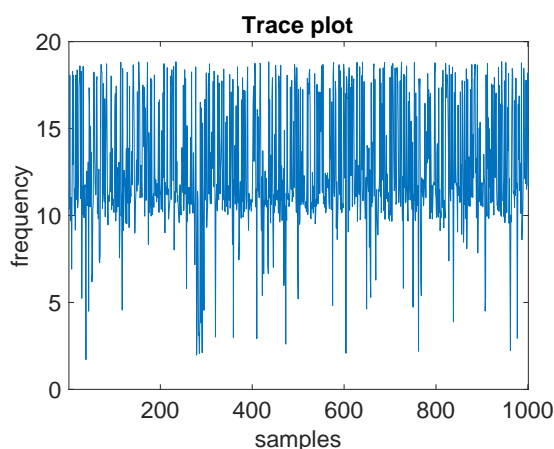


Figure 8. Trace plot of frequency from a pcn-MCMC chain.

- [8] Iglesias M, Park M and Tretyakov M 2018 *Inverse Problems* **34** 105002
- [9] Beskos A, Jasra A, Muzaffer E and Stuart A 2015 *Statistics and Computing* **25** 727–737
- [10] Beskos A, Jasra A, Law K, Marzouk Y and Zhou Y 2018 *SIAM/ASA Journal on Uncertainty Quantification* **6** 762–786
- [11] Jonas L, Papaioannou I and Ullmann E 2018 *Journal of Computational Physics* **368** 154–178
- [12] Reich S 2013 *SIAM Journal on Scientific Computing* **35** A2013–A2024
- [13] Reich S and Cotter C 2015 *Probabilistic forecasting and Bayesian data assimilation* (Cambridge University Press)
- [14] Ruchi S and Dubinkina S 2018 *Nonlin. Processes Geophys.* **25** 731–746
- [15] Chada N K, Iglesias M A, Roininen L and Stuart A M 2018 *Inverse Problems* **34** 055009
- [16] Iglesias M 2016 *Inverse Problems* **32** 025002
- [17] Bear J 1972 *Dynamics of Fluids in Porous Media* (New York: Dover Publications)
- [18] Carrera J and Neuman S 1986 *Water Resources Research* **22(2)** 228–242
- [19] Hanke M 1997 *Inverse Problems* **13** 79–95
- [20] Stuart A 2010 Inverse problems: a Bayesian perspective *Acta Numerica* vol 19 pp 451 – 559
- [21] Matérn B 1986 *Spatial Variation* (Lecture Notes in Statistics, No. 36, Springer)
- [22] Jasra A, Stephens D, Doucet A and Tsagaris T 2011 *Scandinavian Journal of Statistics* **38** 1–22
- [23] Acevedo W, de Wiljes J and Reich S 2017 *SIAM Journal on Scientific Computing* **39** A1834–A1850
- [24] Pele O and Werman M 2009 Fast and robust earth mover’s distances *Computer vision, 2009 IEEE 12th international conference on (IEEE)* pp 460–467
- [25] Cuturi M 2013 Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in Neural Information Processing Systems* (Lake Tahoe, Nevada) pp 2292–2300
- [26] Lehtinen M, Paivarinta L and Somersalo E 1989 *Inverse Problems* **5** 599
- [27] Chustagulprom N, Reich S and Reinhardt M 2016 *SIAM/ASA Journal on Uncertainty Quantification* **4** 592–608
- [28] Cheng Y and Reich S 2015 *Assimilating data into scientific models: An optimal coupling perspective* (Cham: Springer International Publishing) pp 75–118 ISBN 978-3-319-18347-3

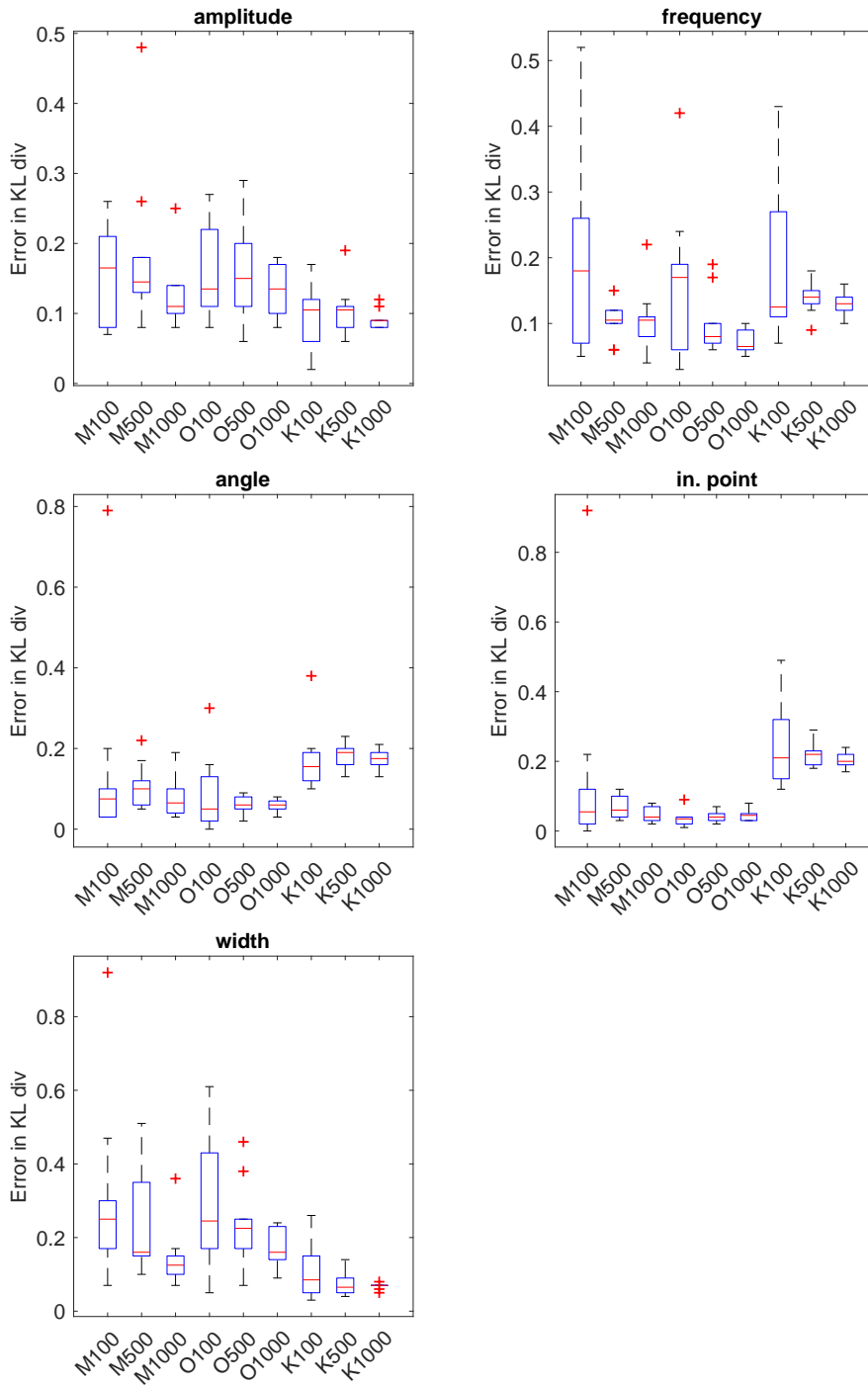


Figure 9. Box plot of KL divergence for geometrical parameters for P2 inference. Central mark is the median, edges of the box are the 25th and 75th percentiles, whiskers extend to the most extreme datapoints, and crosses are outliers. On x-axis numbers stand for ensemble sizes, M stands for monomial based SMC, O for optimal transport based SMC, and K for EKI.

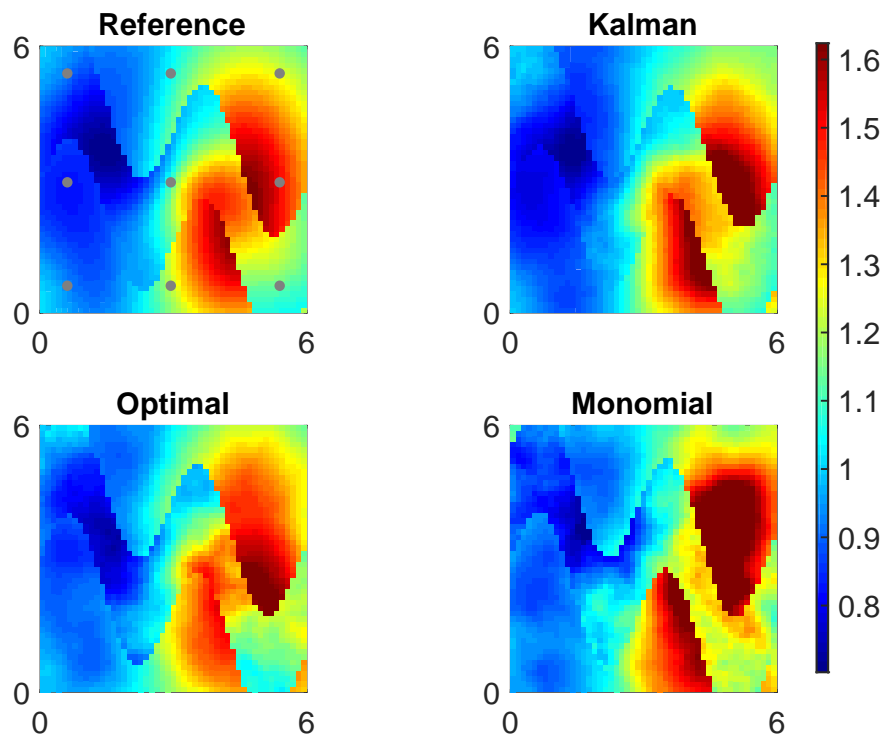


Figure 10. Mean permeability for P2 inference for the lowest error at ensemble size 1000. Observation locations are shown in circles.