CrossMark

Sequential Monte Carlo methods for Bayesian elliptic inverse problems

Alexandros Beskos[1] · **Ajay Jasra**[2] · **Ege A. Muzaffer**[2] · **Andrew M. Stuart**[3]

Abstract In this article, we consider a Bayesian inverse problem associated to elliptic partial differential equations in two and three dimensions. This class of inverse problems is important in applications such as hydrology, but the complexity of the link function between unknown field and measurements can make it difficult to draw inference from the associated posterior. We prove that for this inverse problem a basic sequential Monte Carlo (SMC) method has a Monte Carlo rate of convergence with constants which are independent of the dimension of the discretization of the problem; indeed convergence of the SMC method is established in a function space setting. We also develop an enhancement of the SMC methods for inverse problems which were introduced in Kantas et al. (SIAM/ASA J Uncertain Quantif 2:464–489, 2014); the enhancement is designed to deal with the additional complexity of this elliptic inverse problem. The efficacy of the methodology and its desirable theoretical properties, are demonstrated for numerical examples in both two and three dimensions.

✉ Andrew M. Stuart
   a.m.stuart@warwick.ac.uk

   Alexandros Beskos
   alex@stats.ucl.ac.uk

   Ajay Jasra
   staja@nus.edu.sg

   Ege A. Muzaffer
   m.ege85@nus.edu.sg

[1] Department of Statistical Science, University College London, London WC1E 7HB, UK

[2] Department of Statistics & Applied Probability, National University of Singapore, Singapore 117546, Singapore

[3] Department of Mathematics, University of Warwick, Coventry CV4 7AL, UK

1 Introduction

The viability of the Bayesian approach to inverse problems was established in the pioneering text (Kaipio and Somersalo 2005) which, in particular, demonstrated the potential for Markov chain Monte Carlo (MCMC) methods in this context. Nonetheless, the high-dimensional nature of the unknown, often found from discretizing a field, leads to difficult problems in the design of proposals which are cheap to implement, yet which mix efficiently. One recent approach to tackle these problems has been the development of algorithms with mesh-free mixing times, such as those highlighted in Cotter et al. (2013), Hoang et al. (2013); these non-standard MCMC algorithms avoid the unnecessary penalties incurred by naive proposals related to exploration of the part of the parameter space dominated by the prior. Nonetheless, in the large dataset or small observational noise regimes, one is still confronted with an inference problem in high dimensions which, whilst of smaller order than the dimension of the PDE solver, exhibits wide variations in scales in different coordinates of the parameterizations, leading to substantial challenges for algorithmic tuning.

A different approach, which we will adopt here, involves SMC samplers (Del Moral et al. 2006). These are particle methods which, in the context of Bayesian inverse problems, build an approximation to a sequence of measures which interpolate from the prior to the posterior; the sequential nature of the approximation allows for adaptation of the particle distribution and weights from the (typically simple) prior to the (potentially very complex) posterior. Recent work in the context of inverse problems

(Kantas et al. 2014) has shown how, using the afore-mentioned dimension-independent MCMC methods within SMC, it is possible to construct algorithms which combine the desirable dimension-independent aspects of novel MCMC algorithms with the desirable self-adaptation of particle methods. This combination is beneficial for complex posteriors such as those arising in the large dataset or small noise regimes; in particular the computational results in Kantas et al. (2014) demonstrate an order of magnitude speed-up of these new SMC methods over the MCMC methods highlighted in Cotter et al. (2013), within the context of the inverse problem for the initial condition of the Navier–Stokes equation. Furthermore, recent works (Beskos et al. 2014a; Del Moral 2004; Del Moral et al. 2006; Jasra et al. 2011) have shown that important aspects of this SMC algorithm for inverse problems, such as adaptation, tempering and parallelization, have the potential to provide effective methods even for high-dimensional inverse problems.

The contributions of this article are three-fold:

(1) A computational study of SMC methods for a class of Bayesian inverse problems which arise in applications such as hydrology (Iglesias et al. 2013), and are more challenging to fit, in comparison to the Navier–Stokes inverse problem which was the focus of the development of novel SMC methods in Kantas et al. (2014); furthermore, with modification of the measurement set-up, the inverse problems considered also find application in medical imaging problems such as EIT (Kaipio and Somersalo 2005).

(2) An enhancement of the class of SMC methods introduced in Kantas et al. (2014) which leads to greater efficiency and, in particular, the ability to efficiently solve the elliptic inverse problems which are the focus of this paper.

(3) A proof of the fact that these SMC algorithms have Monte Carlo convergence rates which are mesh-independent and indeed converge in the function space setting. This complements related theoretical work (Hairer et al. 2014) which establishes mesh-independence mixing for the novel MCMC methods which are used as proposal kernels within the SMC approach of Kantas et al. (2014) which we build upon here.

This article is structured as follows. In Sect. 2, we describe the Bayesian model and associated PDE. In Sect. 3 our computational procedure is outlined, along with our theoretical results. In Sect. 4, we present our numerical results. The article is concluded in Sect. 5 with a discussion of areas for future work.

2 Modelling

Consider two normed linear spaces, $\mathcal{K}$ and $\mathcal{Y} \subseteq \mathbb{R}$, corresponding to the state-space of the parameters ($u \in \mathcal{K}$) and observations ($y \in \mathcal{Y}$), respectively. We will observe data at spatial locations $x \in \mathcal{X} \subset \mathbb{R}^d$, $d \in \{2, 3\}$ and we denote the observation at location x as y_x. Let $G : \mathcal{X} \times \mathcal{K} \to \mathcal{Y}$ and, for each $x \in \mathcal{X}$, let $\epsilon_x \in \mathcal{Y}$ be a random variable of zero mean; then we will be concerned with models of the form:

$$y_x = G(x; u) + \epsilon_x.$$

Here $G(x; u)$ is an underlying system behaviour for a given parameter u, and ϵ_x expresses measurement (and indeed sometimes model error) at location x. In our context, G is associated to the solution of a PDE, with parameter u. We are interested in drawing inference on u, given a prior distribution on $u \in \mathcal{K}$, conditional upon observing realizations of y_x for a set of points $x \in O \subseteq \mathcal{X}$, with $\mathrm{Card}(O) < +\infty$. This is the framework of our Bayesian inverse problem. In Sect. 2.1 we define the forward model, and in Sect. 2.2, we describe prior modelling on our unknown. Section 2.3 shows that the posterior distribution is well-defined and states a key property of the log-likelihood, used in what follows.

2.1 Forward model

In this paper, we focus on the general scenario where the forward map G is described by an elliptic PDE. In particular, we work with a problem of central significance in hydrology, namely the estimation of subsurface flow from measurements of the pressure (hydraulic head) at certain locations x in the domain of interest. The pressure and velocity are linked by Darcy's law in which the subsurface permeability appears as a parameter; estimating it is thus a key step in predicting the subsurface flow. In this subsection, we define the forward map from permability to pressure space.

In detail, we consider the d-dimensional cube $\mathcal{X} = [\frac{-\pi}{2}, \frac{\pi}{2}]^d$ as our domain, in both the cases $d = 2, 3$. Define a mapping $p : \mathcal{X} \to \mathbb{R}$, denoting pressure (or hydraulic head), $v : \mathcal{X} \to \mathbb{R}^3$, denoting a quantity proportional to velocity, and $u : \mathcal{X} \to \mathbb{R}$, denoting permeability (or hydraulic conductivity) of soil (McLaughlin and Townley 1996). The behaviour of the system is described through the elliptic PDE:

$$v(x) = -u(x)\nabla_x p(x), \quad x \in \mathcal{X} \tag{1a}$$

$$-\nabla_x \cdot \big(v(x)\big) = f(x), \quad x \in \mathcal{X} \tag{1b}$$

$$p(x) = 0, \quad x \in \partial\mathcal{X}. \tag{1c}$$

Equation (1a) is Darcy's law and contains the permeability u (for us the key parameter); Eq. (1b) expresses continuity of

mass and here $f : \mathcal{X} \to \mathbb{R}$ is assumed to be known and characterizes the source/sink configuration; in Eq. (1c) $\partial \mathcal{X}$ is the boundary of the domain, and we are thus assuming a homogeneous boundary condition on the boundary pressure—other boundary conditions, specifying the flux, are also possible. Together Eq. (1) define an elliptic PDE for pressure p.

If u is in $L^\infty(\mathcal{X})$ and lower bounded by a positive constant $k_{\min}$ a.e. in $\mathcal{X}$ then, for every $f \in H^{-1}(\mathcal{X})$, there is a unique solution $p \in H_0^1(\mathcal{X})$ to the PDE (1) satisfying

$$\|p\|_{H_0^1} \leq \frac{1}{k_{\min}} \|f\|_{H^{-1}}; \tag{2}$$

see Dashti and Stuart (2015) and the references therein. In this setting, the forward map $G(x; u) := p(x)$ is well-defined and thus corresponds to solution of the elliptic PDE for a given permeability field u. A typical choice of the source/sink function f is

$$f(x) = \sum_i c_i \delta_{x_i}(x). \tag{3}$$

The set of points $\{x_i\}$ denotes the known position of sources or sinks, and the signs of each c_i determine whether one has a source or sink at x_i (Iglesias et al. 2013). We note that the cleanest setting for the mathematical formulation of the problem requires $f \in H^{-1}(\mathcal{X})$ and, in theory, will require mollification of the Dirac's at each x_i; in practice this modification makes little difference to the inference.

2.2 Prior modelling of permeablity u

We describe the modelling of u in three dimensions; simplification to the two-dimensional setting is straightforward. We begin by expressing the unknown model parameter as a Fourier series:

$$u(x) = \bar{u}(x) + \sum_{k \in \mathbb{Z}^3} u_k e_k(x). \tag{4}$$

Here we have scaled Fourier coefficients

$$e_k(x) = a_k \exp(ik \cdot x), \quad k \in \mathbb{Z}^3 \tag{5}$$

and the real coefficients $\{a_k\}$, complex coefficients $\{u_k\}$ (satisfying $\bar{u}_k = u_{-k}$) and real-valued $L^\infty(\mathcal{X})$ function $\bar{u}$ will be chosen to enforce the mathematically necessary (and physically sensible) positivity restriction

$$u(x) \geq k_{\min} > 0, \qquad x \in \mathcal{X}. \tag{6}$$

The use of Fourier series in principle enables the representation of arbitrary functions in $L^2(\mathcal{X})$ by use of periodic extensions. However we will impose a rate of decay on the $\{a_k\}$, in order to work in the setting of inversion for this problem, as developed in Hoang et al. (2013), Schwab and

Stuart (2012); this rate of decay will imply a certain degree of smoothness in the function $(u - \bar{u})(\cdot)$. Noting that the functions $\exp(ik \cdot x)$ have $L^\infty(\mathcal{X})$ norm equal to one, we can place ourselves in the setting of Hoang et al. (2013), Schwab and Stuart (2012) by assuming that, for some $q > 0$, $C > 0$,

$$\sum_k |a_k|_\infty < \infty, \quad \sum_{k:|k|_\infty > j} |a_k|_\infty < Cj^{-q}. \tag{7}$$

We choose a_k of the form

$$a_k = a|k|_{L^\infty}^{-\alpha} \tag{8}$$

and then impose $\alpha > 3$ in dimension $d = 3$ or $\alpha > 2$ in dimension $d = 2$.

Given this set-up, we need to find a suitable prior for u, so that the forward model $G(x; u)$ is almost-surely well defined, as well as reflecting any prior statistical information we may have. There are several widely adopted approaches in the literature for prior parameterization of the permeability, the most common being the log-normal choice (see Tarantola (2005) for details and, for example, Iglesias et al. (2013) for a recent application), widely adopted by geophysicists, and the uniform case Hoang et al. (2013), Schwab and Stuart (2012) which has been successfully adopted in the computational mathematics literature, building on earlier work of Schwab in uncertainty quantification (Schwab and Gittelson 2011). We work with the uniform priors popularized by Schwab: we choose $u_k \overset{i.i.d.}{\sim} \mathcal{U}_{[-1,1]}$ in the representation (4) for u, resulting in a pushforward measure ν_0 on u as in Schwab and Stuart (2012). We let $\mathcal{K}$ denote the separable Banach space found from the closure, with respect to the L^∞ norm, of the set of functions used in the representation (4) of $u(\cdot)$. Then ν_0 is viewed as a measure on $\mathcal{K}$; see Dashti and Stuart (2015) for further details. Once the parameters a_k are chosen to satisfy (7), the mean function $\bar{u}(x)$ can be chosen to ensure that there is $k_{\min}$ such that $u(\cdot)$ is in $L^\infty(\mathcal{X})$ and satisfies (6) almost surely with respect to the prior ν_0 on function u.

2.3 Bayesian inverse problem

We observe the pressure at certain locations, the set of which is denoted as $O \in \mathcal{X}$. We will suppose that for each $x \in O$ and independently, $\epsilon_x \sim \mathcal{N}(0, \sigma^2)$, where $\mathcal{N}(0, \sigma^2)$ is the normal distribution of mean 0 and known variance σ^2. Then the log-likelihood is, up to an irrelevant additive constant, given by

$$\Phi(u; y) := -\sum_{x \in O} \frac{|G(x; u) - y(x)|^2}{2\sigma^2}. \tag{9}$$

Along with the prior modelling in Sect. 2.2, this defines a scenario so that the forward model $G(x; \cdot)$ is, almost-surely, well defined and, in fact, Lipschitz. As in Dashti and Stuart (2015), Schwab and Stuart (2012) we may then define a

posterior ν^y on u which has density with respect to ν_0 given by (9)

$$\frac{d\nu^y}{d\nu_0}(u) \propto \pi(u) = \exp\big(-\Phi(u; y)\big). \qquad (10)$$

Exploring the posterior distribution ν^y is the objective of the paper. In doing so, the following fact will be relevant; it is easily established by the fact that (6) holds almost surely for $u \sim \nu_0$, together with the bound on the solution of the elliptic PDE given in (2).

Lemma 2.1 *There is a constant* $\pi_{\min} = \pi_{\min}(y) > 0$ *such that* $\pi_{\min} \leq \pi(u) \leq 1$ *almost surely for* $u \sim \nu_0$.

We finish by noting that, in algorithmic practice, it is typically necessary (see, however, Agapiou et al. (2014) in the context of MCMC) to apply a spectral truncation:

$$u(x) = \bar{u}(x) + \sum_{k:\, |k|_\infty < c} u_k e_k(x), \qquad (11)$$

where c is a truncation parameter. Having defined the desired parameterization of u, we consider the truncated vector of Fourier coefficients u_k as the object of inference in practice.

3 Sequential Monte Carlo

In this section, we describe the application of SMC methods to Bayesian inversion. Sect. 3.1 contains an explanation of the basic methodology and statement of a basic (non-adaptive) algorithm. Sect. 3.2 contains statement and proof of a convergence theorem for the basic form of the algorithm, notably because it applies in infinite-dimensional spaces. In Sect. 3.3, we describe an adaptive version of the SMC algorithm, which we use in practice.

3.1 Standard SMC samplers

Let $(E, \mathcal{E})$ denote a measure space and ν_0 a probability measure on that space. We wish to sample from a target probability measure ν on $(E, \mathcal{E})$, which has density with respect to ν_0 known up to a normalizing constant:

$$\frac{d\nu}{d\nu_0}(u) \propto \pi(u). \qquad (12)$$

We introduce a sequence of "bridging" densities which enable us to connect ν_0 to ν:

$$\pi_n(u) \propto \pi(u)^{\varphi_n}, \quad x \in E, \qquad (13)$$

where $0 = \varphi_0 < \cdots < \varphi_{n-1} < \varphi_n < \cdots < \varphi_p = 1$; we refer to the φ_j as *temperatures*. We let ν_n denote the probability measure with density proportional to π_n with respect

to ν_0. Assuming that $\pi(u)$ is finite ν_0 almost surely we obtain

$$\frac{d\nu_n}{d\nu_0}(u) \propto \pi(u)^{\varphi_n}, \quad \frac{d\nu_n}{d\nu_{n-1}}(u) \propto \ell_{n-1}(u) := \pi(u)^{\varphi_n - \varphi_{n-1}},$$
$$n \in \{1, \ldots, p\}. \qquad (14)$$

We note that the assumption on π being finite is satisfied for our elliptic inverse problem; see Lemma 2.1. Although $\nu = \nu_p$ may be far from ν_0, careful choice of the φ_n can ensure that ν_n is close to ν_{n-1} allowing gradual evolution of approximation of ν_0 into approximation of ν. Other choices of bridging densities are possible and are discussed in e.g. Del Moral et al. (2006).

Let $\{L_{n-1}\}_{n=1}^p$ denote the sequence of (nonlinear) maps on measures found by applying Bayes's Theorem with likelihood proportional to $\{\ell_{n-1}\}_{n=1}^p$ and let $\{K_n\}_{n=1}^p$ be a sequence of Markov kernels (and equivalently, for notational convenience, the resulting linear maps on measures) with invariant measure $\{\nu_n\}_{n=1}^p$. We define $\{\Phi_n\}_{n=1}^p$ to be the nonlinear maps on measures found as $\Phi_n = K_n L_{n-1}$. Explicitly we have, for each $n \geq 1$ and any probability measure μ on E:

$$(\Phi_n \mu)(du) = \frac{\mu(\ell_{n-1} K_n(du))}{\mu(\ell_{n-1})},$$

where we use the notation $\mu(\ell_{n-1} K_n(du)) = \int_E \mu(dy) \ell_{n-1}(y) K_n(y, du)$ and $\mu(\ell_{n-1}) = \int_E \mu(dy) \ell_{n-1}(y)$. It then follows that

$$\nu_n = \Phi_n \nu_{n-1}, \quad n \in \{1, \ldots, p\}. \qquad (15)$$

The standard SMC algorithm is described in Fig. 1. It involves a population of M particles evolving with n. With no resampling, the algorithm coincides with annealed importance sampling as in Neal (2001). With resampling at every step [i.e. the case $M_{\text{thresh}} = M$, where M_{thresh} denotes the cutoff point for the effective sample size (EES)] we define the empirical approximating measures by the iteration

$$\nu_n^M = S^M \Phi_n \nu_{n-1}^M, \quad n \in \{1, \ldots, p\}; \qquad \nu_0^M = \frac{1}{M} \sum_{m=1}^M \delta_{u_0^m}. \qquad (16)$$

Here

$$(S^M \mu)(dv) = \frac{1}{M} \sum_{m=1}^M \delta_{v^{(m)}}(dv), \quad v^{(m)} \sim \mu \ \ \text{i.i.d.}.$$

3.2 Convergence property

The issue of dimensionality in SMC methods has attracted substantial attention in the literature (Beskos et al. 2014a, b, c;

Fig. 1 Standard SMC Samplers. $M_{\text{thres}} \in \{1, \ldots, M\}$ is a user defined parameter

0. Sample $\{u_0^m\}_{m=1}^M$ i.i.d. from ν_0 and define the weights $w_0^m = M^{-1}$ for $m = 1, \cdots, M$.

Set $n = 1$ and $l = 0$.

1. For each m set $\hat{w}_n^m = \ell_{n-1}(u_{n-1}^m)w_{n-1}^m$ and sample u_n^m from $K_n(u_{n-1}^m, \cdot)$; calculate the normalized weights

$$w_n^m = \hat{w}_n^m / \left(\sum_{m=1}^M \hat{w}_n^m \right).$$

2. Calculate the Effective Sample Size (ESS):

$$ESS_{(n)}(M) := \frac{\left(\sum_{m=1}^M w_n^m \right)^2}{\sum_{m=1}^M (w_n^m)^2}. \tag{17}$$

If $ESS_{(n)}(M) \leq M_{\text{thres}}$:

resample $\{u_n^m\}_{m=1}^M$ according to the normalized weights $\{w_n^m\}_{m=1}^M$;

re-initialise the weights by setting $w_n^m = M^{-1}$ for $m = 1, \cdots, M$;

let $\{u_n^m\}_{m=1}^M$ now denote the resampled particles.

3. If $n < p$ set $n = n + 1$ and return to Step 1; otherwise stop.

Rebeschini and Van Handel 2015). In this section, using a simple approach for the analysis of particle filters which is clearly exposed in Rebeschini and Van Handel (2015), we show that for our SMC method it is possible to prove dimension-free error bounds. Whilst the theoretical result in this subsection is not entirely new (similar results follow from the work in Del Moral (2004), Del Moral (2013)), a direct and simple proof is included for non-specialists in SMC methods, to highlight the utility of SMC methods in inverse problems, and to connect with related recent results in dimension-independent MCMC, such as Hairer et al. (2014), which are far harder to establish.

We will consider the algorithm in Fig. 1 with $M_{\text{thres}} = M$, so one resamples at every time step (and this is multinomially). Note that then, for $n \geq 0$, at the end of each step of the algorithm the approximation to ν_n is given by

$$\nu_n^M(du) := \frac{1}{M} \sum_{m=1}^M \delta_{u_n^m}(dx), \tag{18}$$

which follows from the algorithm in Fig. 1 with $M_{\text{thres}} = M$ or, equivalently, (16).

Throughout, we will assume that there exists a $\kappa > 0$ such that for each $n \geq 0$ and any $u \in E$

$$\kappa \leq \ell_n(u) \leq 1/\kappa. \tag{19}$$

We note that this holds for the elliptic inverse problem from the previous section, when the uniform prior ν_0 is employed; see Lemma 2.1.

Let P denote the collection of all probability measures on E. Let $\mu = \mu(\omega)$ and $\nu = \nu(\omega)$ denote two possibly random elements in P, and $\mathbb{E}^\omega$ expectation w.r.t. ω. We define the distance between $\mu, \nu \in \mathsf{P}$ by

$$d(\mu, \nu) = \sup_{|f|_\infty \leq 1} \sqrt{\mathbb{E}^\omega |\mu(f) - \nu(f)|^2},$$

where the supremum is over all $f : E \to \mathbb{R}$ with $|f|_\infty := \sup_{v \in E} |f(v)| \leq 1$. This definition of distance is indeed a metric on the space of random probability measures; in particular it satisfies the triangle inequality. In the context of SMC, the randomness underlying the approximations (18) comes from the various sampling operations within the algorithm.

We have the following convergence result for the SMC algorithm.

Theorem 3.1 *Assume* (19) *and consider the SMC algorithm with $M_{\text{thresh}} = M$. Then, for any $n \geq 0$,*

$$d(\nu_n^M, \nu_n) \leq \sum_{j=0}^n (2\kappa^{-2})^j \frac{1}{\sqrt{M}}.$$

Proof For $n = 0$ the result holds via Lemma 3.2. For $n > 0$, we have, by the triangle inequality, Lemma 3.1 and Lemma 3.2 (which may be used by the conditional independence structure of the algorithm),

$$d(\nu_n, \nu_n^M) = d(\Phi_n \nu_{n-1}, S^M \Phi_n \nu_{n-1}^M)$$
$$\leq d(\Phi_n \nu_{n-1}, \Phi_n \nu_{n-1}^M)$$

$$+ d(\Phi_n v_{n-1}^M, S^M \Phi_n v_{n-1}^M)$$
$$\leq \frac{2}{\kappa^2} d(v_{n-1}, v_{n-1}^M) + \frac{1}{\sqrt{M}}.$$

Iterating gives the desired result. □

Remark 3.1 This theorem shows that the sequential particle filter actually reproduces the true posterior distribution v_p, in the limit $M \to \infty$. We make some comments about this.

- The measure v_p is well-approximated by v_p^M in the sense that, as the number of particles $M \to \infty$, the approximating measure converges to the true measure. The result holds in the infinite-dimensional setting. As a consequence, the algorithm as stated is robust to finite dimensional approximation.
- In principle, the theory applies even if the Markov kernel K_n is simply the identity mapping on probability measures. However, moving the particles according to a non-trivial v_n-invariant measure is absolutely essential for the methodology to work in practice. This can be seen by noting that if K_n is indeed taken to be the identity map on measures then the particle positions will be unchanged as n changes, meaning that the measure v_p is approximated by weighted samples (almost) from the prior, clearly undesirable in general.
- The MCMC methods in Cotter et al. (2013) provide explicit examples of Markov kernels with the desired property of preserving the measures v_n, including the infinite-dimensional setting.
- In fact, if the Markov kernel K_n has some ergodicity properties then it is sometimes possible to obtain bounds which are *uniform* in p; see Del Moral (2004), Del Moral (2013).

Lemma 3.1 *Assume* (19). *Then, for any* $n \geq 1$ *and any* $\mu, v \in \mathsf{P}$,

$$d(\Phi_n \mu, \Phi_n v) \leq \frac{2}{\kappa^2} d(\mu, v).$$

Proof For any measurable $f : E \to \mathbb{R}$ we have

$$[\Phi_n \mu - \Phi_n v](f) = \frac{1}{\mu(\ell_{n-1})}[\mu - v](\ell_{n-1} K_n(f))$$
$$+ \frac{v(\ell_{n-1} K_n(f))}{\mu(\ell_{n-1}) v(\ell_{n-1})}[v - \mu](\ell_{n-1}).$$

So we have, by Minkowski,

$$\mathbb{E}^\omega[|[\Phi_n \mu - \Phi_n v](f)|^2]^{1/2}$$
$$\leq \mathbb{E}^\omega\left[\left|\frac{1}{\mu(\ell_{n-1})}[\mu - v](\ell_{n-1} K_n(f))\right|^2\right]^{1/2}$$
$$+ \mathbb{E}^\omega\left[\left|\frac{v(\ell_{n-1} K_n(f))}{\mu(\ell_{n-1}) v(\ell_{n-1})}[v - \mu](\ell_{n-1})\right|^2\right]^{1/2}.$$

Note that the ratio

$$\frac{v(\ell_{n-1} K_n(f))}{v(\ell_{n-1})}$$

is an expectation of f and is hence bounded by 1 in modulus, if $|f|_\infty \leq 1$. Then using the fact that $|\ell_{n-1} K_n(f)|_\infty \leq \kappa^{-1}$ and $\ell_{n-1} \geq \kappa$ (see (19)) we deduce that

$$\mathbb{E}^\omega[|[\Phi_n \mu - \Phi_n v](f)|^2]^{1/2}$$
$$\leq \frac{1}{\kappa^2}\mathbb{E}^\omega\left[\left|[\mu - v](\ell_{n-1} K_n(f)\kappa)\right|^2\right]^{1/2}$$
$$+ \frac{1}{\kappa^2}\mathbb{E}^\omega\left[\left|[v - \mu](\ell_{n-1}\kappa)]\right|^2\right]^{1/2}.$$

using the fact that $|\ell_{n-1} K_n(f)|_\infty \leq \kappa^{-1}$ and $\ell_{n-1} \leq \kappa^{-1}$, with the first following from (19) together with the Markov property for K_n, taking suprema over f completes the proof. □

Lemma 3.2 *The sampling operator satisfies*

$$\sup_{\mu \in \mathsf{P}} d(S^M \mu, \mu) \leq \frac{1}{\sqrt{M}}.$$

Proof Let v be an element of $\mathsf{P}(X)$ and $\{v^{(k)}\}_{k=1}^M$ a set of i.i.d. samples with $v^{(1)} \sim v$; the randomness entering the probability measures is through these samples, expectation with respect to which we denote by $\mathbb{E}^\omega$ in what follows. Then

$$S^M v(f) = \frac{1}{M} \sum_{k=1}^M f(v^{(k)})$$

and, defining $\overline{f} = f - v(f)$, we deduce that

$$S^M v(f) - v(f) = \frac{1}{M} \sum_{k=1}^M \overline{f}(v^{(k)}).$$

It is straightforward to see that

$$\mathbb{E}^\omega \overline{f}(v^{(k)})\overline{f}(v^{(l)}) = \delta_{kl}\mathbb{E}^\omega|\overline{f}(v^{(k)})|^2.$$

Furthermore, for $|f|_\infty \leq 1$,

$$\mathbb{E}^\omega|\overline{f}(v^{(1)})|^2 = \mathbb{E}^\omega|f(v^{(1)})|^2 - |\mathbb{E}^\omega f(v^{(1)})|^2 \leq 1.$$

It follows that, for $|f|_\infty \leq 1$,

$$\mathbb{E}^\omega|v(f) - S^M v(f)|^2 = \frac{1}{M^2} \sum_{k=1}^M \mathbb{E}^\omega|\overline{f}(v^{(k)})|^2 \leq \frac{1}{M}.$$

Since the result is independent of v, we may take the supremum over all probability measures and obtain the desired result. □

3.3 Adaptive-SMC samplers

In practice, the SMC samplers algorithm requires the specification of $0 < \varphi_0 < \cdots < \varphi_{n-1} < \varphi_n < \cdots < \varphi_p = 1$ as well as any parameters in the MCMC kernels. As demonstrated in Jasra et al. (2011), Kantas et al. (2014), the theoretical validity of which is established in Beskos et al. (2014d), these parameters may be set on the fly.

First, we focus on the specification of the sequence of distributions. Given step $n - 1$ and $\pi_{n-1}(x)$, we select the next target density by adapting the temperatures to a required value for the effective sample size (ESS) statistic (17) as in Jasra et al. (2011) (see also Zhou et al. (2015) for an alternative procedure). So, for a user-specified threshold M_{thres}, we choose φ_n as the solution of $ESS_{(n)}(M) = M_{\text{thres}}$. One can use an inexpensive bisection method to obtain φ_n.

Second, we turn to the specification of the mutation kernels K_n. Several options are available here, but we will use reflective random-walk Metropolis proposals on each univariate component, conditionally independently. We will adapt the random-move proposal scales, $\epsilon_{j,n}$, with j the co-ordinate and n the time index. A simple choice would be to tune $\epsilon_{j,n}$ to the marginal variance along the j-th co-ordinate; since this is analytically unavailable we opt for the SMC estimate at the previous time-step. Thus, we set $\epsilon_{n,j} = \rho_n \sqrt{\widehat{\text{Var}}(U_{n,j})}$ where ρ_n is a global scaling parameter. For ρ_n itself, we propose to modify it based on the previous average acceptance rate over the population of particles (denoted α_{n-1}^N), to try to have average acceptance rates in a neighbourhood of 0.2 (see e.g. Beskos et al. (2009) and the references therein for a justification). Our adaptive strategy works as follows;

$$\rho_n = \begin{cases} 2\rho_{n-1}, & \text{if } \alpha_{n-1}^N > 0.3 \\ 0.5\rho_{n-1}, & \text{if } \alpha_{n-1}^N < 0.15 \\ \rho_{n-1}, & \text{o.w.} \end{cases} \tag{20}$$

Thus, we scale ρ upwards (downwards) if the last average acceptance rate went above (below) a predetermined neighbourhood of 0.2. This approach is different to the work in Kantas et al. (2014).

In addition, one can synthesize a number, say M_n, of baseline MCMC kernels to obtain an overall effective one with good mixing; this is a new contribution relative to Kantas et al. (2014). To adapt M_n, we follow the following heuristic; We propose to select M_n using $M_n = \lfloor \frac{m}{\rho_n^2} \rfloor$, with m being a global parameter. The intuition is that for random-walk-type transitions of increment with small standard deviation δ, one needs $\mathcal{O}(\delta^{-2})$ steps to travel distance $\mathcal{O}(1)$ in the state-space. A final modification for practical computational reasons is that we force M_n steps to lie within a predetermined bounded set, i.e. $[l, u]$.

The adaptive-SMC algorithm works as in Fig. 1, except in step 1, before simulation from K_n is undertaken, our adaptive procedure is implemented. Then one may resample (or not) and then move the samples according to K_n. In addition, the algorithm will run for a random number of time steps and terminate when $\varphi_n = 1$ (which will happen in a finite number of steps almost surely).

4 Numerical results

In this section, we describe the details of our implementation (Sect. 4.1), describe the objects of inference (Sect. 4.2) and give our results in 2D (Sect. 4.3) and 3D (Sect. 4.4).

4.1 Implementation details

The software used in our experiments has been implemented in C++ for the GNU\Linux platform. We used the Libmesh library for finite elements computation (Kirk et al. 2006), we used the Fast Fourier Transform for rapid evaluation of the sum in $u(\cdot)$ at predetermined grid-points in $\mathcal{X}$ and we exploited parallel computation wherever possible, for which we used the MPI libraries. Our experiments were run on a computer server with 23 "Intel(R) Xeon(R)CPU X7460 @2.66GHz" processors, each with 2 cores; 50 Gb memory and running "RedHat Linux version 2.6.18-194.el5" operating system. The experiments discussed in this paper used 20 processors.

All the colour plots of random fields (e.g. permeability fields) have been prepared using the rainbow colour scheme from the R programming language/environment. The scheme quantizes the Hue quantity of HSV (Hue Saturation Value) triplet of a pixel. Our level of quantization is selected to be 256 (8 bits), with the Hue range of $[0, 1]$, hence we normalize the random fields to this range and quantize to 8 bits to get the Hue value for a pixel. Saturation and Value were taken to be 1. All images were computed using 500×500 equi-spaced point evaluations from the respective random fields.

4.2 Objects of inference

The work in Völlmer (2013) investigates the performance of the Bayesian approach for our elliptic inverse problem and gives sufficient conditions under which posterior consistency holds. Posterior consistency is concerned with "ball probabilities" of type

$$\lim_{\text{Card}(O) \to \infty} \int_{B_\epsilon} \frac{d\nu^y}{d\nu_0}(u)\nu_0(du) = 1,$$

where $y = \{y_x\}_{x \in O}$ and B_ϵ is the ϵ neighbourhood of the true value of u. One way to check such a result numerically

Table 1 Parameter values of used for the 2D experiments

Parameter name	Value
Frequency cutoff	10
Finite elements d.o.f.	100
σ^2	5×10^{-7}
M	1000
M_{thres}	600
a	4
$\bar{u}$	40
Wall-clock time	11 h

Between 5 and 1000 steps are allowed for the iterates of the MCMC kernels. The frequency cutoff determines the level of discretization of the permeability field. Finite elements d.o.f. denotes the number of finite elements used in the numerical solution of the elliptic PDE, higher values indicate better approximation at the expense of computational resources. For a see (8)

is to use the posterior estimates obtained via our method. The estimated ball probabilities are computed as follows:
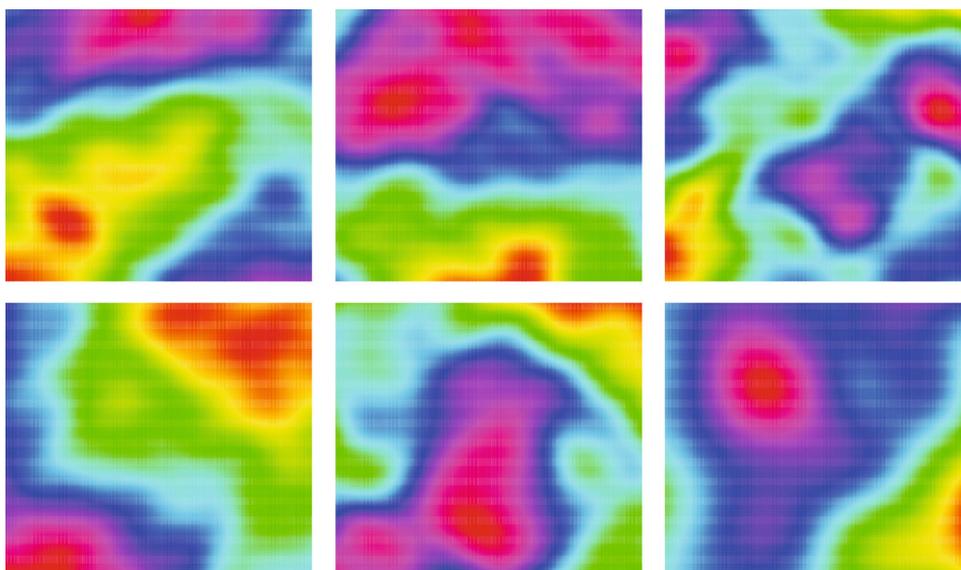
$$\sum_i w_p^i \mathbb{I}_{B_\epsilon}(u(x_p^i)) \tag{21}$$

Although not all the conditions in Völlmer (2013) required for posterior consistency to hold are fulfilled, we will nonetheless empirically investigate such a consistency property. This also provides a severe test for the SMC method since it implies posterior measures in the large dataset limit.

4.3 Results in 2D

We consider Eq. (1) in dimension $d = 2$ and with source and sinks as specified in (3). Our goal is to construct a sequence of posterior estimates, corresponding to increasing number of observations in order to numerically illustrate posterior consistency. Table 1 shows the parameters used in our experiments.

To get an empirical sense of these parameters' effects on the distribution of the permeability field, we plot some samples from the prior field $u(x)$ in Fig. 2. We will then generate 100 data points from the model, in the scenario where the permeability field is as Fig. 3a. In Fig. 3b we show the mean of the posterior permeability field, using these 100 noisily observed data points. The posterior mean is estimated as the mean of the particle distribution from SMC at the final time, and notable in that it shows good agreement with true value of the permeability.

In another experiment, designed to study posterior consistency, a sequence of posterior estimates is formed by repeatedly running the adaptive SMC algorithm with, respectively, 4, 16, 36, 64 and 100 observations equi-spaced inside the domain of $[-\pi/2, \pi/2]^2$. The computed MSE and ball probabilities are given in Fig. 4, with the ball radius ϵ taken to be 0.17×360, where 360 is the number of parameters in the system, corresponding to a frequency cutoff of 10. The Figure suggests that as more data become available posterior consistency is obtained as predicted, under slightly more restrictive assumptions than we have in play here, in Völlmer (2013). This is interesting for two reasons: firstly it suggests the potential for more refined Bayesian posterior consistency analyses for nonlinear PDE inverse problems; secondly it demonstrates the potential to solve hard practical Bayesian inverse problems and to obtain informed inference from a relatively small number of observations.

Finally, Fig. 5 shows marginal posterior density estimates corresponding to 144 observations. The usual observation is

Fig. 2 Six permeability field samples drawn from the prior

Fig. 3 The true and posterior estimated permeability field. The estimated filed is the mean estimated at the final step of the SMC method. **a** True permeability field. **b** Estimated permeability field
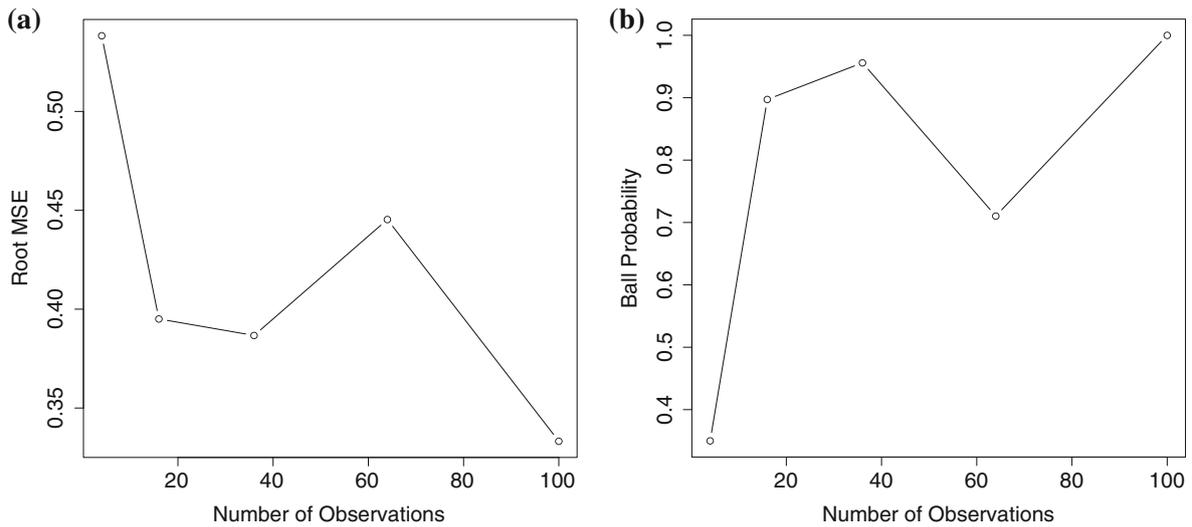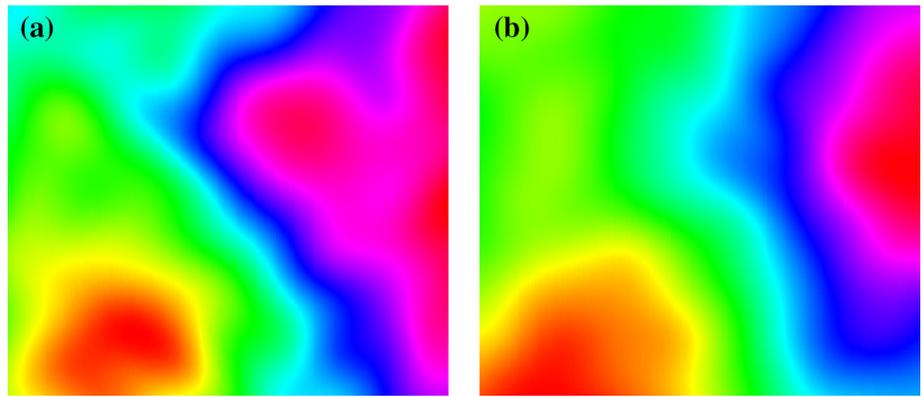

Fig. 4 Numerical consistency checks for the sequence of experiments with 4, 16, 36, 64 and 100 observations. **a** Root mean squared error corresponding to the sequence of experiments. **b** Estimated small-ball probabilities (Eq. 21) corresponding to the sequence of experiments
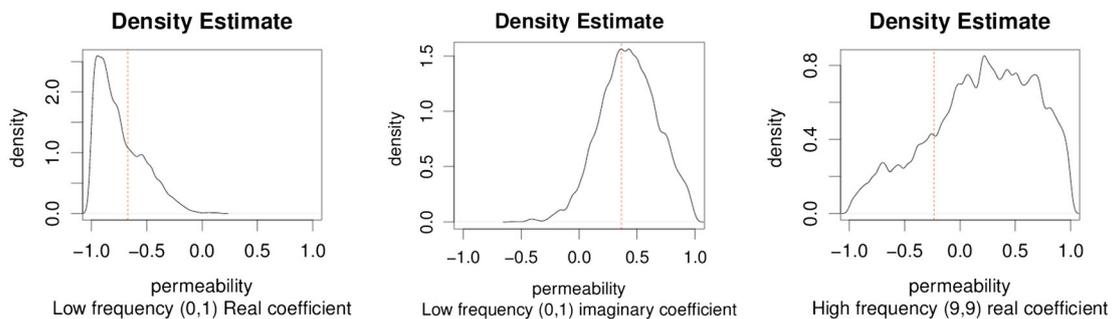

Fig. 5 Posterior marginal density estimates for two low and one high frequency coefficients in the 2D case

to note the effectiveness of even the mode estimator in lower frequencies. Another important observation is the similarity of the high frequency marginal densities to the prior. In fact, it is this behaviour that makes a prior invariant MCMC proposal superior to others, i.e. the proposal itself is almost optimal for a wide range of coefficients in the problem.

4.4 Results in 3D

A more realistic experiment is performed using the 3D setup discussed in Sect. 2. In this setup, the computational aspects of the problem are further highlighted as the numerical solution of the forward operator becomes much harder due to the increased cardinality of the finite elements basis. The values

Table 2 Parameter values used for the 3D experiment discussed in this section

Parameter name	Value
# of observations	125
Frequency cutoff	5
Finite elements d.o.f.	1000
σ^2	1×10^{-8}
M	1000
M_{thres}	600
a	1
$\bar{u}$	100
Wall-clock time	10 days

Between 5 and 200 steps are allowed for the iterates of the MCMC kernels

of parameters in this numerical study are given in Table 2. The data are generated from the model, under the specifications given in Table 2.

In Fig. 6, we consider the performance of our SMC algorithm in this very challenging scenario. In Fig. 6a, we can see the average acceptance rates of the MCMC moves, over the time parameter of the SMC algorithm. We can observe that these acceptance rates do not collapse to zero and are not too far from 0.2. This indicates that the step-sizes are chosen quite reasonably by the adaptive SMC algorithm and the MCMC kernels have some mixing ability. In Fig. 6b, we can see the number of MCMC iterations that are used per-particle over the time parameter of the SMC algorithm. We can observe, as one might expect, that as the target distribution becomes more challenging, the number of MCMC steps required grows. Figure 6 indicates reasonable performance of our SMC algorithm.

In terms of inference, the posterior density estimates are shown in Fig. 7. Recall that the priors are uniform. These estimates indicate a clear deviation from the prior specification, illustrating that the data influence our inference significantly. This is not obvious, and establishes that one can hope to use this Bayesian model in real applications.

5 Summary

In this article, we have presented an SMC method for Bayesian inverse problems and applied it to a particular ellip-
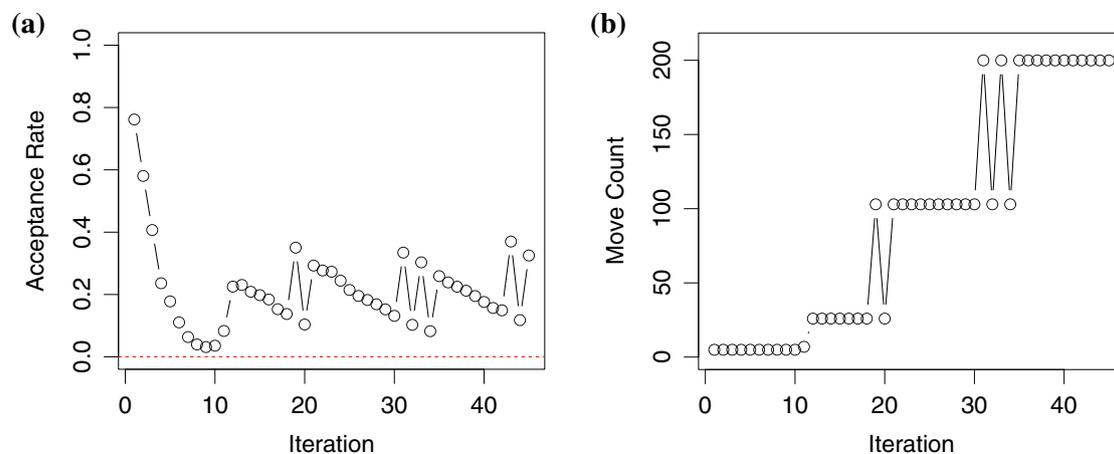

Fig. 6 SMC performance for 3D example. **a** Average acceptance rates for each iteration of the adaptive-SMC algorithm. **b** Adaptively selected MCMC steps for each iteration of the adaptive-SMC algorithm, bounded to be in [5,200]
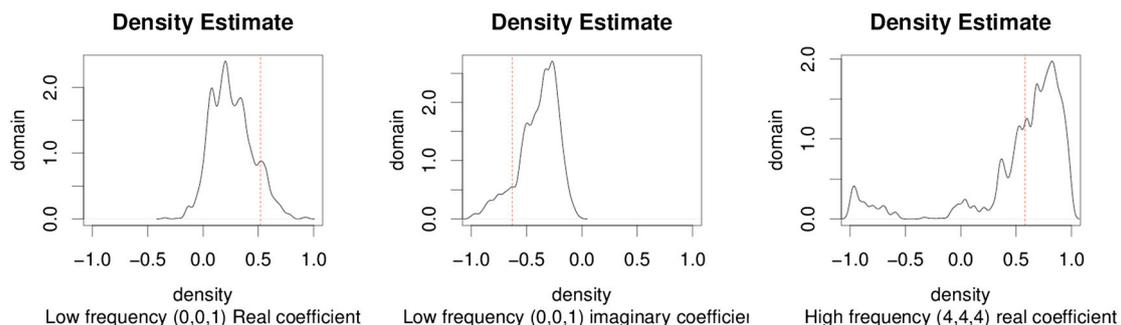

Fig. 7 Posterior marginal density estimates for two low and one high frequency coefficients in the 3D case

tic PDE inversion; the methodology, however, is transferable to other PDE inverse problems. Simulations demonstrated both the feasibility of the SMC method for challenging infinite-dimensional inversion, as well as the property of posterior contraction to the truth. In addition to simulations, we have provided a straightforward proof of the fact that SMC methods are robust to the dimension of the problem.

There are several avenues for future research. Firstly, our error bounds explode w.r.t. the time parameter. It is of interest to find realistic conditions for which this is not the case (for instance the bounds in Del Moral (2004), Del Moral (2013), Whiteley (2013) have assumptions which either do not hold or are hard to verify). Secondly, a further algorithmic innovation is to use multi-level Monte Carlo method as in Hoang et al. (2013), within the SMC context; this is being considered in Beskos et al. (2015). And finally, it is of interest to consider the use of these methods to solve other Bayesian inference problems.

References

Agapiou, S., Roberts, G.O., Völlmer, S.: Unbiased Monte Carlo: posterior estimation for intractable/infinite-dimensional models (2014). arXiv:1411.7713

Beskos, A., Roberts, G., Stuart, A.M.: Optimal scalings for local Metropolis-Hastings chains on non-product targets in high-dimensions. Ann. Appl. Probab. **19**, 863–898 (2009)

Beskos, A., Crisan, D., Jasra, A.: On the stability of sequential Monte Carlo methods in high dimensions. Ann. Appl. Probab. **24**, 1396–1445 (2014a)

Beskos, A., Crisan, D., Jasra, A., Whiteley, N.P.: Error bounds and normalizing constants for sequential Monte Carlo samplers in high-dimensions. Adv. Appl. Probab. **46**, 279–306 (2014b)

Beskos, A., Crisan, D., Jasra, A., Kamatani, K., Zhou, Y.: A stable particle filter in high-dimensions (2014c). arXiv:1412.3501

Beskos, A., Jasra, A., Kantas, N., Thiery, A.: On the convergence of adaptive sequential Monte Carlo methods (2014d). arxiv:1306.6462

Beskos, A., Jasra, A., Law, K.J.H., Tempone, R., Zhou, Y.: Multilevel sequential Monte Carlo samplers (2015). arXiv preprint

Cotter, S., Roberts, G., Stuart, A.M., White, D.: MCMC methods for functions: modifying old algorithms to make them faster. Stat. Sci. **28**, 424–446 (2013)

Dashti, M., Stuart, A.M.: The Bayesian approach to inverse problems. In: Handbook of Uncertainty Quantification. Springer (2015). arxiv:1302.6989

Del Moral, P.: Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications. Springer, New York (2004)

Del Moral, P., Doucet, A., Jasra, A.: Sequential Monte Carlo samplers. J. R. Stat. Soc. B **68**, 411–436 (2006)

Del Moral, P.: Mean Field Simulation for Monte Carlo Integration. Chapman & Hall, London (2013)

Hairer, M., Stuart, A.M., Vollmer, S.J.: Spectral gaps for a Metropolis-Hastings algorithm in infinite dimensions. Ann. Appl. Probab. **24**, 2455–2490 (2014)

Hoang, V.H., Schwab, C., Stuart, A.M.: Complexity analysis of accelerated MCMC methods for Bayesian inversion. Inverse Prob. **29**, 085010 (2013)

Iglesias, M., Law, K.J.H., Stuart, A.M.: Evaluation of Gaussian approximations for data assimilation for reservoir models. Comput. Geosci. **17**, 851–885 (2013)

Jasra, A., Stephens, D.A., Doucet, A., Tsagaris, T.: Inference for Lévy driven stochastic volatility models using adaptive sequential Monte Carlo. Scand. J. Stat. **38**, 1–22 (2011)

Kaipio, J., Somersalo, E.: Statistical and Computational Inverse Problems. Applied Mathematical Sciences, vol. 160. Springer, New York (2005)

Kantas, N., Beskos, A., Jasra, A.: Sequential Monte Carlo for inverse problems: a case study for the Navier Stokes equation. SIAM/ASA J. Uncertain. Quantif. **2**, 464–489 (2014)

Kirk, B.S., Peterson, J.W., Stogner, R.H., Carey, G.F.: A C++ library for parallel adaptive mesh refinement/coarsening simulations. Eng. Comput. **22**, 237–254 (2006)

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

Neal, R.M.: Annealed importance sampling. Stat. Comput. **11**, 125–139 (2001)

Rebeschini, P., Van Handel, R.: Can local particle filters beat the curse of dimensionality? Ann. Appl. Probab. (to appear)

Schwab, C., Gittelson, C.J.: Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. Acta Numer. **20**, 291–467 (2011)

Schwab, C., Stuart, A.M.: Sparse determinisitc approximation of Bayesian inverse problems. Inverse Prob. **28**, 045003 (2012)

Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM, Philadelphia (2005)

Völlmer, S.: Posterior Consistency for Bayesian inverse problems through stability and regression results. Inverse Prob. **29**, 125011 (2013)

Whiteley, N.: Sequential Monte Carlo samplers: Error bounds and insensitivity to initial conditions. Stoch. Anal. Appl. **30**, 774–798 (2013)

Zhou, Y., Johansen, A.M., Aston, J.A.D.: Towards automatic model comparison: an adaptive sequential Monte Carlo approach. J. Comput. Graph. Stat. (to appear)