Dynamic updating of numerical model discrepancy using sequential sampling

2014 Inverse Problems 30 114019

(http://iopscience.iop.org/0266-5611/30/11/114019)

IOP Publishing Inverse Problems
Inverse Problems **30** (2014) 114019 (19pp) doi:10.1088/0266-5611/30/11/114019

Dynamic updating of numerical model discrepancy using sequential sampling

Daniela Calvetti[1], **Oliver Ernst**[2] **and Erkki Somersalo**[1]

[1] Case Western Reserve University, Department of Mathematics, Applied Mathematics and Statistics, 10900 Euclid Ave., Cleveland, OH 44106, USA
[2] Case Western Reserve University, Department of Physics, 10900 Euclid Ave., Cleveland, OH 44106, USA

E-mail: dxc57@case.edu, oke@case.edu and ejs49@case.edu

Received 16 April 2014, revised 22 May 2014
Accepted for publication 27 May 2014
Published 28 October 2014

Abstract

This article addresses the problem of compensating for discretization errors in inverse problems based on partial differential equation models. Multi-dimensional inverse problems are by nature computationally intensive, and a key challenge in practical applications is to reduce the computing time. In particular, a reduction by coarse discretization of the forward model is commonly used. Coarse discretization, however, introduces a numerical model discrepancy, which may become the predominant part of the noise, particularly when the data is collected with high accuracy. In the Bayesian framework, the discretization error has been addressed by treating it as a random variable and using the prior density of the unknown to estimate off-line its probability density, which is then used to modify the likelihood. In this article, the problem is revisited in the context of an iterative scheme similar to Ensemble Kalman Filtering (EnKF), in which the modeling error statistics is updated sequentially based on the current ensemble estimate of the unknown quantity. Hence, the algorithm learns about the modeling error while concomitantly updating the information about the unknown, leading to a reduction of the posterior variance.

Keywords: modeling error, ensemble Kalman filter, sampling

(Some figures may appear in colour only in the online journal)

1. Introduction

In this paper, we consider the inverse problem of estimating a distributed parameter, modeled as a non-constant coefficient function of a partial differential equation (PDE) from data

comprising of limited and noisy observations of the solution of the equation. The electrical impedance tomography (EIT) problem provides a representative example of such a problem [5]. Since controlled ground truth corresponding to measured data is often difficult to obtained, computational methods are invariably validated with synthetic, numerically generated data. It is a commonly accepted practice in inverse problems that inverse solvers should not be based on the same forward solver that is used to generate the test data, to avoid what is known as an 'inverse crime'. Results based on inverse crime data tend to be overly optimistic, because the synthetic data contains only artificially added noise that is typically independent of the unknown parameter of interest. In reality, the noise model should account also for the model discrepancy, that is, the difference between the model and the reality. Synthetic data, while never capturing all the intricacies of reality, should at least provide the challenge of not being generated with a model identical to the one that the inverse solver is based on.

In statistics literature, the insufficiency of a model to describe the reality is referred to as model discrepancy, and finding ways to account for it is a topic of active study, see, e.g., [2, 12, 17]. The model discrepancy due to numerical discretization error has been addressed in the literature in the context of Bayesian inverse problems, [1, 16], together with other types of model mismatches such as poorly known geometry [19, 20], boundary clutter in the data [3], or inadequate modeling of the underlying physics [10, 11, 24].

In this article, we address the problem of accounting for the numerical discretization error. We consider this problem in the context of EIT, where the PDE model is assumed to provide a good approximation of the physical measurement setting, while the discretization of the equation, e.g., by the finite element method (FEM), generates an approximation error. The model discrepancy grows with the coarseness of the mesh, and consequently a fine mesh is desirable to decrease the error. However, finer meshing increases the computational complexity of the inverse problem, and therefore the selection of the mesh density becomes a trade-off between complexity and accuracy.

The numerical discretization error depends on the underlying unknown conductivity distribution, which makes the analysis of the model discrepancy a challenge. It has been pointed out in literature, however, that while in the deterministic framework this dependency may be a bottleneck for error modeling, in the Bayesian statistical framework a very natural solution exists: the unknown conductivity is modeled as a random variable with a known prior probability density, and therefore the modeling error also has a distribution that can be computationally estimated. This idea leads to a statistical model for the discrepancy, and accounting for it in the inverse solver has been demonstrated to provide a significant improvement in the quality of the computed solution. This is true in particular when the signal-to-noise ratio is high, and the modeling error becomes the dominant component of the noise.

In [1, 16], the modeling error was estimated by an off-line calculation that is performed prior to solving the inverse problem. More precisely, a representative sample from the prior density of the unknown of primary interest is drawn, and the probability density of the modeling error is estimated by computing the model predictions using two discretization meshes, one representing an accurate approximation of the continuous model, the other being the discretization mesh used in the inverse solver. In this work, we investigate the possibility of estimating the modeling error distribution in a dynamic fashion, by updating the estimate while we iteratively learn about the unknown of primary interest. In particular, unlike in the cited works, the estimate of the modeling error distribution is not based on draws from the prior but rather on forward simulations using the current ensemble of estimates, analogously to the classical estimation of the error of numerical solvers for differential equations. The proposed algorithm bears a similarity with Ensemble Kalman Filtering (EnKF) algorithm,

with an artificial time that refers to the updating round. EnKF in the framework for time-invariant inverse problems was recently proposed and investigated in [13]. Unlike in the cited article, in the present article the time evolution is related to the updating of the error model. The idea of estimating the modeling error dynamically rather than off-line has been previously proposed in [6], in which the authors use a reduced order surrogate model in delayed acceptance Markov Chain Monte–Carlo (MCMC) computations, updating the modeling error statistics as the sampler moves on. The present article can be seen as a fast approximate version of the full-scale posterior MCMC posterior sampling of the modeling error. We demonstrate by a computed example that the dynamic modeling error update leads to reduced posterior variance of the unknown.

2. Discretization error and model discrepancy

In this section, we start by reviewing the discretization error model as discussed in [1, 15, 16], and we arrive at a formulation of the forward model. We then present the new sequential approach to the problem, and finally apply the method to the model problem of EIT.

2.1. Discretization error model

Consider a linear model with an unknown parameter,

$$\mathsf{A}_\theta x = y, \tag{1}$$

where y is a given input, θ is a parameter vector, x is the model output, and A_θ is a parameter-dependent matrix, which we assume to be square and invertible. We restrict our attention to a finite dimensional estimation problem, the particular example that we have in mind being a discretized version of a linear elliptic PDE with unknown coefficients, θ representing the degrees of freedom of the unknown. The vector y is the source term, x is the discretized solution of the PDE, while the matrix A_θ represents the FEM or finite difference (FD) matrix corresponding to the model.

Assume that the data depend linearly on the solution x of the above system, such as boundary values of the solution. The observation model, including the observation noise, is written as

$$b = \mathsf{B}x + e = \mathsf{B}\mathsf{A}_\theta^{-1} y + e = f(\theta) + e, \tag{2}$$

where f is the non linear mapping from the unknown to the presumably noiseless data, and e is the additive noise vector. To set up the inverse problem in the Bayesian framework, the noise e is modeled as a random variable. For simplicity, we assume that the additive noise is zero mean Gaussian, $e \sim \mathcal{N}(0, \mathsf{C})$, leading to a likelihood model

$$b|\theta \sim \mathcal{N}(f(\theta), \mathsf{C}).$$

Here, for the time being, it is assumed that the model (2) corresponds to a discretization fine enough so that is can be considered as an accurate approximation of the continuous model, and e accounts only for the exogenous measurement noise. We model the unknown θ *a priori* as a Gaussian random variable with mean θ_0 and covariance Γ, leading to a prior model

$$\theta \sim \mathcal{N}\big(\theta_0, \Gamma\big). \tag{3}$$

The posterior density is then given by

$$\pi(\theta|b) \propto \exp\left(-\frac{1}{2} \| b - f(\theta) \|_{\mathsf{C}}^2 - \frac{1}{2} \| \theta - \theta_0 \|_{\Gamma}^2 \right),$$

where we use the norms $\| z \|_{\mathsf{C}}^2 = z^{\mathsf{T}} \mathsf{C}^{-1} z$, $\| z \|_{\Gamma}^2 = z^{\mathsf{T}} \Gamma^{-1} z$. In particular, the maximum *a posteriori* (MAP) estimate of θ, denoted by $\widehat{\theta}$, satisfies

$$\widehat{\theta} = \operatorname{argmin} \left\{ \| b - f(\theta) \|_{\mathsf{C}}^2 + \| \theta - \theta_0 \|_{\Gamma}^2 \right\}.$$

Ideally, if computational resources and computing time are not restricted, we may increase the accuracy of the discretization in the model (1) to make the observation model (2) fine enough so that the discretization error can be neglected, as in the likelihood model above. For inverse problems, however, this is often not a viable assumption, since the computation of estimates such as $\widehat{\theta}$ may be expensive due to the non linearity of the forward mapping f, and fine discretization leads to increased computational burden. Also, the computation time per estimate may be restricted by the data acquisition scheme. Therefore, it is not uncommon to resort to a rather coarse discretization in the inverse solver, introducing possible a significant approximation error not accounted for by the noise term e.

To analyze the approximation error, in the inverse solver, consider two discretization levels of the forward model, denoted by

$$\mathsf{A}_\theta^N x = y^N, \quad \mathsf{A}_\theta^n x = y^n, \tag{4}$$

where $\mathsf{A}^N \in \mathbb{R}^{N\times N}$, $\mathsf{A}^n \in \mathbb{R}^{n\times n}$, $n < N$, and n corresponds to a coarse model, while N refers to a finely discretized model. The numbers n and N can be assumed to represent the number of nodes in the FEM or FD meshes. It is assumed here that the fine discretization is such that, within a prescribed numerical accuracy, it can be considered as a sufficient approximation of the continuous model. We assume that the forward problem is coercive, thus the matrices A^N and A^n are invertible and we consider the corresponding noiseless observation models,

$$b^N = \mathsf{B}^N \left(\mathsf{A}_\theta^N\right)^{-1} y^N = f^N(\theta), \quad b^n = \mathsf{B}^n \left(\mathsf{A}_\theta^n\right)^{-1} y^n = f^n(\theta), \tag{5}$$

where $\mathsf{B}^N \in \mathbb{R}^{m\times N}$, $\mathsf{B}^n \in \mathbb{R}^{m\times n}$ are the linear observation models for the corresponding discretized problems. The idea of the Bayesian discretization error modeling can be summarized as follows: assuming that the fine grid model corresponds with satisfactory accuracy to the measurement, while the coarse grid model is the surrogate used in the inverse problem, we may write the noisy observation model as

$$b = f^N(\theta) + e = f^n(\theta) + \left\{ f^N(\theta) - f^n(\theta) \right\} + e,$$

where the modeling error vector, defined as

$$m = F^{N,n}(\theta) = f^N(\theta) - f^n(\theta),$$

is modeled as a random variable. If π_θ denotes the probability density of θ, then the density of m is obtained as the push-forward density,

$$\pi_m = F_*^{N,n} \pi_\theta,$$

that is,

$$\mathsf{P}\{m \in A\} = \int_{\left(F^{N,n}\right)^{-1}(A)} \pi_\theta(\theta)\, \mathrm{d}\theta = \int_A F_*^{N,n} \pi_\theta(m)\, \mathrm{d}m.$$

Since the modeling error depends on the unknown θ, our information about the error relies on the current information about θ, and estimating the probability density of the modeling error becomes part of the Bayesian inverse problem.

In [15, 16], it was proposed to use the prior probability density, $\pi_\theta = \mathcal{N}(\theta_0, \Gamma)$, to estimate the modeling error distribution, which can carried out off-line. In practice, a Gaussian approximation of the modeling error distribution was proposed, leading to the following algorithm for computing the MAP estimate.

Algorithm 1. Sampling-based modeling error approximation

(i) Generate a sample of realizations of θ,

$$\{\theta_1, \theta_2, \ldots, \theta_K\},$$

drawing them from the prior distribution $\mathcal{N}(\theta_0, \Gamma)$.

(ii) Compute a sample of model error vectors,

$$m_\ell = F^{N,n}(\theta_\ell), \quad 1 \leqslant \ell \leqslant K,$$

and further, the sample mean and covariance,

$$\overline{m} = \frac{1}{K}\sum_{\ell=1}^{K} m_\ell, \quad \Sigma = \frac{1}{K}\sum_{\ell=1}^{K} (m_\ell - \overline{m})(m_\ell - \overline{m})^{\mathsf{T}}.$$

(iii) Compute the MAP estimate for θ using the updated enhanced error model,

$$b = f^n(\theta) + E, \quad E \sim \mathcal{N}(\overline{m}, \mathsf{C} + \Sigma),$$

where it is assumed that E is independent of θ.

Observe that the updated error model can be used for approximation of the whole posterior density, not just for the MAP estimate. The above algorithm neglects the mutual correspondence of the unknown θ and the modeling error. Nonetheless, its practical viability has been demonstrated in numerous applications, see, e.g., [1, 4, 18]. From the point of view of computational efficiency, the costly part of this algorithm is step (ii), requiring a repeated solution of the fine grid forward model. However, this computation needs to be done only once and off-line; once the model error mean and covariance are available, they can be used repeatedly for solving the inverse problem. The algorithm is therefore particularly attractive in applications in which the inverse problem needs to be solved repeatedly and in real time, such as in electrical impedance process tomography [14, 18]. For preliminary results concerning the effect of the interdependency of m and θ on the error modeling, we refer to [3, 16]. In this work, the question is elaborated further.

The central idea of the dynamic estimation method in this context is the following: the prior density of θ allows a preliminary estimate of the modeling error distribution, which is then used to update the density of θ. As the information about the unknown increases, so does the information about the modeling error. Following this observation, we therefore propose a method of dynamically updating the distributions of θ and m. The implementation of this alternating algorithm is inspired by EnKF, in which the artificial time evolution is related to the update of the likelihood, taking the latest information of the unknown and thus the implied information of modeling error into account. As the EnKF algorithm is known to work well with a small sample size, the alternating algorithm remains computationally light and fast. The iterative nature of the approach also makes it possible to monitor the convergence of the

solution, allowing to stop the iterations as the desired quality of the estimation is met. Moreover, the algorithm is ideal for adaptation to situations in which the model evolves in time.

2.2. Sampling through randomized optimization

The alternating updating scheme requires sampling of the parameter θ from the current approximation of the posterior density. In this work, we use a randomized optimization scheme similar to the analysis step in the EnKF algorithm [9], bearing also a similarity to the classical parametric bootstrapping [7, 8]. The randomized sampling has been suggested as an approximate posterior sampling scheme, see [21, 22].

Assume that we have a sample of size k of realizations of the parameter vectors drawn from the prior distribution (19) by independent sampling,

$$\mathcal{S}_k = \{\theta_1, \theta_2, \ldots, \theta_k\}.$$

Assuming that the likelihood is based on the preliminary observation model ignoring the modeling error,

$$b|\theta \sim \mathcal{N}\big(f^n(\theta), \mathsf{C}\big),$$

we first generate an artificial data sample through parametric bootstrapping [7, 8],

$$b_j = b + w_j, \quad w_j \sim \mathcal{N}(0, \mathsf{C}), \quad 1 \leqslant j \leqslant k,$$

where b is the measured data. From this sample and $\mathcal{S}_k$, we construct a family of approximations of the vector θ using the coarse grid model, solving

$$\theta_j^+ = \operatorname{argmin}\Big\{ \| b_j - f^n(\theta) \|_{\mathsf{C}}^2 + \| \theta - \theta_j \|_{\mathsf{\Gamma}}^2 \Big\}.$$

The algorithm provides an updating scheme of the sample,

$$\mathcal{S}_k \to \mathcal{S}_k^+ = \{\theta_1^+, \theta_2^+, \ldots, \theta_k^+\}, \tag{6}$$

The new sample, although not exactly sampled from the posterior density, represents a surrogate model for it; we refer to the literature for the discussion of this approximation [21].

2.3. Iterative update of modeling error

When comparing iterative and direct methods, one attractive feature of the former class is that the iterations can be stopped at any point, and the current estimate is an approximation of the solution. In contrast, if a direct method is terminated early, all computations will go to waste. When applying this logic to the modeling error, from the point of view of the solution to the inverse problem, even a low rank approximation of the modeling covariance may provide a sufficient correction to the measurement noise covariance. However, if the modeling error is estimated off-line, a good stopping criterion may be hard to define *a priori*. We therefore propose a dynamic updating of the modeling error estimate.

The proposed algorithm proceeds by alternating the following two updating steps:

(i) Given a sample of parameters θ, compute an ensemble-based estimate of the density $\pi_m = F_*^{N,n}\pi_\theta$;
(ii) Using the current estimate of π_m in the likelihood, update the sample of the parameter θ, representing the updated density π_θ.

More precisely, let $\mathcal{S}_k^0$ be the initial sample of random draws $\theta_j^0, 1 \leqslant j \leqslant k$ from the prior density $\pi_\theta(\theta) = \pi_{\text{prior}}(\theta)$, and assume that we have a sample updating scheme such as (6),

$$\mathcal{S}_k^t \to \mathcal{S}_k^{t+1}, \quad t = 0, 1, 2, \ldots$$

We denote the elements of $\mathcal{S}_k^t$ by θ_j^t, $1 \leqslant j \leqslant k$. After T steps, we have collected the cumulative sample

$$\mathcal{S}^{(T)} = \bigcup_{t=0}^{T} \mathcal{S}_k^t.$$

To estimate the modeling error from the sample $\mathcal{S}^{(T)}$, we define the modeling error vectors

$$m_j^t = F^{N,n}\left(\theta_j^t\right), \quad 1 \leqslant j \leqslant k, \quad 0 \leqslant t \leqslant T,$$

and compute the cumulative error mean and covariance,

$$\overline{m}^{(T)} = \frac{1}{(T+1)\,k} \sum_{t=0}^{T} \sum_{j=1}^{k} m_j^t,$$

$$\Sigma^{(T)} = \frac{1}{(T+1)\,k} \sum_{t=0}^{T} \sum_{j=1}^{k} \left(m_j^t - \overline{m}^{(T)}\right)\left(m_j^t - \overline{m}^{(T)}\right)^{\mathsf{T}}.$$

Given the ensemble mean and covariance of the sample at $t = T + 1$,

$$\overline{m}^{T+1} = \frac{1}{k} \sum_{j=1}^{k} m_j^{T+1}, \quad \Sigma^{T+1} = \frac{1}{k} \sum_{j=1}^{k} \left(m_j^{T+1} - \overline{m}^{T+1}\right)\left(m_j^{T+1} - \overline{m}^{T+1}\right)^{\mathsf{T}},$$

it is straightforward to verify the recursive formulas for updating the cumulative mean and covariance,

$$\overline{m}^{(T+1)} = \frac{T}{T+1}\overline{m}^{(T)} + \frac{1}{T+1}\overline{m}^{T+1}, \tag{7}$$

and

$$\Sigma^{(T+1)} = \frac{T}{T+1}\Sigma^{(T)} + \frac{1}{T+1}\Sigma^{T+1} + \frac{T}{(T+1)^2}\left(\overline{m}^{(T)} - \overline{m}^{T+1}\right)\left(\overline{m}^{(T)} - \overline{m}^{T+1}\right)^{\mathsf{T}}. \tag{8}$$

With these notations, we can formulate an iterative algorithm based on randomized MAP estimation, combined with an iterative modeling error update scheme.

Algorithm 2. Dynamic modeling error update

(i) Initialize: draw a sample of size k from the prior density,

$$\mathcal{S}_k^0 = \left\{\theta_1^0, \ldots, \theta_k^0\right\}, \quad \theta_j^0 \sim \pi_{\text{prior}}.$$

Set $\overline{m} = 0 \in \mathbb{R}^m$, $\Sigma = 0 \in \mathbb{R}^{m\times m}$. Set $t = 0$.

(ii) Generate a bootstrapped data set, using the current likelihood model,

$$b_j = b - \overline{m} + w_j, \quad w_j \sim \mathcal{N}(0, \mathsf{C} + \Sigma), \quad 1 \leqslant j \leqslant k.$$

(iii) Update the sample,

$$\mathcal{S}_k^{t+1} = \left\{\theta_1^{t+1}, \ldots, \theta_k^{t+1}\right\},$$

where

$$\theta_j^{t+1} = \text{argmin}\left\{ \| b_j - f^n(\theta) \|_{\mathsf{C}+\Sigma}^2 + \| \theta - \theta_j^t \|_{\Gamma}^2 \right\}, \quad 1 \leqslant j \leqslant k.$$

(iv) Compute the new model error sample,

$$m_j^{t+1} = F^{N,n}\left(\theta_j^{t+1}\right), \quad 1 \leqslant j \leqslant k,$$

and the corresponding mean and covariance,

$$\overline{m}^{t+1} = \frac{1}{k}\sum_{j=1}^{k} m_j^{t+1}, \quad \Sigma^{t+1} = \frac{1}{k}\sum_{j=1}^{k}\left(m_j^{t+1} - \overline{m}^{t+1}\right)\left(m_j^{t+1} - \overline{m}^{t+1}\right)^{\mathsf{T}}.$$

(v) Update the cumulative model error mean and covariance,

$$\overline{m}^{+} = \frac{t}{t+1}\overline{m} + \frac{1}{t+1}\overline{m}^{t+1},$$

$$\Sigma^{+} = \frac{t}{t+1}\Sigma + \frac{1}{t+1}\Sigma^{t+1} + \frac{t}{(t+1)^2}\left(\overline{m} - \overline{m}^{t+1}\right)\left(\overline{m} - \overline{m}^{t+1}\right)^{\mathsf{T}}.$$

(vi) If the model error mean and covariance satisfy the convergence criterion, stop; else, increase t by one, set $\overline{m} = \overline{m}^{+}$ and $\Sigma = \Sigma^{+}$, and continue from step (ii).

Before illustrating the performance of algorithm 2 with computed examples, some comments are in order. To keep the computations simple, the algorithm is based on the Gaussian approximation of the modeling error similarly to algorithm 1, and within each iteration, the error is treated as if it were independent of θ. To keep the computational burden light, the sample size k needs to be small: the effect of varying the sample size will be addressed when we discuss computed examples. The modeling error statistics are based on the accumulated parameter sample, while the current estimate for the parameter is the sample mean over the current sample $\mathcal{S}_k^t$, similar to EnKF algorithms. This choice is based on the consideration that samples of realizations of the unknown, even based on draws of θ from the prior, provide useful information about the modeling error. Conversely, the first estimates of θ are usually strongly corrupted by the modeling error that is not yet properly estimated. These observations reflect the ill-posed nature of inverse problems: forward simulations are not sensitive to errors, while inverse solutions are. The stopping criterion may be based either on convergence of the estimate of θ, the convergence of the model error, or on limitations on computational cost.

3. Application to EIT

We apply the proposed algorithm to the problem of EIT. The computations are done by using a two-dimensional model with finite element discretization.

3.1. Forward model

Consider a bounded domain $\Omega \subset \mathbb{R}^2$ with connected smooth boundary $\partial\Omega$, and a conductivity distribution σ: $\Omega \to \mathbb{R}_+$, which is assumed to be a bounded function. The electric voltage potential u satisfies the conservation of charge equation,

$$\nabla \cdot (\sigma \nabla u) = 0 \quad \text{in } \Omega. \tag{9}$$

To specify the boundary condition, we assume that L contact electrodes are attached to the boundary, modeled as disjoint, relatively open and connected subsets $e_\ell \subset \partial\Omega$, $\bar{e}_\ell \cap \bar{e}_j = \varnothing$ for $\ell \neq j$. On each electrode e_ℓ, a constant contact impedance $z_\ell > 0$ is given. The input data consist of a current vector $J \in \mathbb{R}^L$ specifying the electric currents injected through each electrode,

$$J_\ell = \int_{e_\ell} \sigma \frac{\partial u}{\partial n} \mathrm{d}S, \quad 1 \leqslant \ell \leqslant L, \tag{10}$$

the conservation of charge imposing Kirchhoff's law,

$$\sum_{\ell=1}^{L} J_\ell = 0. \tag{11}$$

The current in and out of the body between the electrodes is assumed to vanish, that is,

$$\sigma \frac{\partial u}{\partial n} \Big|_{\partial\Omega \setminus \cup e_\ell} = 0. \tag{12}$$

The current injection induces electric voltages U_ℓ on the electrodes, and they are related to the voltage potential through the model

$$\left(u + z_\ell \sigma \frac{\partial u}{\partial n} \right) \Big|_{e_\ell} = U_\ell, \tag{13}$$

where the parameters $z_\ell > 0$ are the presumably known contact impedances of the electrodes.

The system (9)–(13) can be written in a variational form,

$$\begin{aligned} \mathcal{B}((v, V), (u, U)) &= \int_\Omega \sigma \nabla v \cdot \nabla u \mathrm{d}x + \sum_{\ell=1}^{L} \frac{1}{z_\ell} \int_{e_\ell} \big(v - V_\ell\big)\big(u - U_\ell\big) \mathrm{d}S \\ &= \sum_{\ell=1}^{L} V_\ell J_\ell, \end{aligned} \tag{14}$$

and the quadratic form $\mathcal{B}$ can be shown to be coercive in $\big(H^1(\Omega) \times \mathbb{R}^L\big)/\mathbb{R}$, that is, the voltage potential is in the Sobolev space of smoothness index one, $U \in \mathbb{R}^L$, and the soultion is unique up to the selection of the ground. The ambiguity of an additive constant in the solution pair (u, U) can be removed by specifying the ground condition, which is chosen here by imposing the condition

$$\sum_{\ell=1}^{L} U_\ell = 0. \tag{15}$$

For further details, we refer to [23].

3.2. Discretized EIT model

To discretize the problem, we need to specify two finite element meshes, a fine one with N nodes, and a coarse one with $n < N$ nodes. Figure 1 shows the meshes used with $L = 16$ electrodes on the boundary of a disk-like domain. The conductivity is discretized by representing it as a piecewise constant function in the coarse mesh,

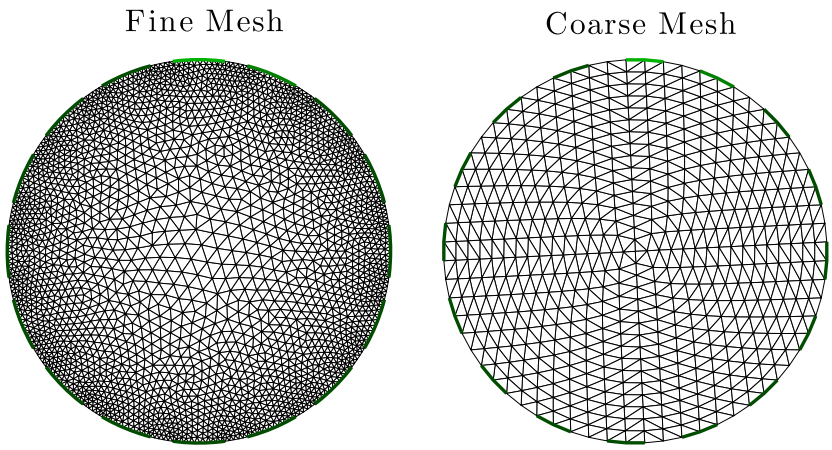

Figure 1. The fine grid (left), consisting of $N_e = 4579$ elements, and the coarse grid (right) with $n_e = 1024$ elements. The number of nodes are $N = 2675$ and $n = 545$, respectively.

$$\sigma(x) = \sigma_0 + \sum_{j=1}^{n_e} \theta_j \chi_j^n(x), \tag{16}$$

where χ_j^n is the characteristic function of the jth triangular element of the coarse grid, the number of elements in the coarse mesh being n_e. Throughout this work, we keep the background conductivity σ_0 fixed.

Consider first the discretization in the coarse mesh. We approximate the voltage potential in a nodal basis,

$$u(x) = \sum_{j=1}^{n} \alpha_j \psi_j(x).$$

To implement the ground condition, we represent the voltage vector in a basis $\left\{U^1, \ldots, U^{L-1}\right\}$, each basis vector satisfying the condition (15), and write

$$U = \sum_{j=1}^{L-1} \beta_j U^j.$$

Using these representations, we write the equation

$$\sum_{j=0}^{n} \alpha_j \mathcal{B}\Big((v, V), \big(\psi_j, 0\big)\Big) + \sum_{j=1}^{L-1} \beta_j \mathcal{B}\Big((v, V), \big(0, U^j\big)\Big) = V^{\mathsf{T}} J,$$

and by choosing the test functions v and V among the basis vectors, we finally arrive at the matrix equation

$$\mathsf{A}_\theta^n x = y^n,$$

where

$$x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \in \mathbb{R}^{n+L-1}, \quad y = \begin{bmatrix} 0 \\ \left[\left(U^j \right)^{\mathrm{T}} J \right]_{j=1:L-1} \end{bmatrix} \in \mathbb{R}^{n+L-1}, \tag{17}$$

and the matrix $\mathsf{A}_\theta^n \in \mathbb{R}^{(n+L-1)\times(n+L-1)}$ has the block structure

$$\mathsf{A}_\theta^n = \begin{bmatrix} \left[\mathcal{B}\Big(\big(\psi_i, 0\big), \big(\psi_j, 0\big) \Big) \right]_{i,j=1:n} & \left[\mathcal{B}\Big(\big(\psi_i, 0\big), \big(0, U^j\big) \Big) \right]_{i=1:n, j=1:L-1} \\ \left[\mathcal{B}\Big(\big(0, U^i\big), \big(\psi_j, 0\big) \Big) \right]_{i=1:L-1, j=1:n} & \left[\mathcal{B}\Big(\big(0, U^i\big), \big(0, U^j\big) \Big) \right]_{i,j=1:L-1} \end{bmatrix}.$$

To define the observation model, we assume that a full frame of electric currents are applied to the body, and the corresponding voltage patterns are measured. A frame consists of the maximum number of linearly independent current patterns satisfying the Kirchhoff condition (11). We denote the chosen frame by $\left\{ J^{(1)}, \ldots, J^{(L-1)} \right\}$. For each current pattern, we denote the corresponding vector y in (17) by $y^{(\ell)}$, and write

$$\begin{bmatrix} \alpha^{(\ell)} \\ \beta^{(\ell)} \end{bmatrix} = \left(\mathsf{A}_\theta^n \right)^{-1} y^{(\ell)}.$$

Further, by defining the basis matrix

$$\mathsf{U} = \left[U^1 \;\; U^2 \;\; \cdots \;\; U^{L-1} \right] \in \mathbb{R}^{L\times(L-1)},$$

the voltage pattern $U^{(\ell)}$ corresponding to the current pattern $J^{(\ell)}$ is given by

$$U^{(\ell)} = \mathsf{U}\beta^{(\ell)} = \mathsf{U}\Big[\mathsf{O}_{(L-1)\times n} \;\; \mathsf{I}_{L-1} \Big] \left(\mathsf{A}_\theta^n \right)^{-1} y^{(\ell)} = \mathsf{B}^n \left(\mathsf{A}_\theta^n \right)^{-1} y^{(\ell)},$$

where I_{L-1} is the $(L-1)\times(L-1)$ unit matrix and $\mathsf{O}_{(L-1)\times n}$ is a zero matrix of indicated size. The full measurement is then obtained by stacking the vectors $U^{(\ell)}$ into a single vector of length $L\,(L-1)$.

To complete the description of the discretization, consider the fine grid discretization, consisting of N nodes. In this case, the FEM model is constructed in the same way as in the coarse grid case, with the only extra feature that the conductivity σ given in (16) needs to be projected on the finer mesh, which can be done, e.g., by defining the value of the conductivity in an element to be the integrated average of the coarse grid conductivity over the element.

3.3. Prior density

We define the prior density in the coarse grid using the following spatial prior model. Let x_j denote the center of mass of the jth coarse grid element, $1 \leqslant j \leqslant n_{\mathrm{e}}$. We define the prior covariance matrix $\Gamma \in \mathbb{R}^{n_{\mathrm{e}}\times n_{\mathrm{e}}}$ as

$$\Gamma_{j\ell} = \gamma \exp\left(-\frac{\left| x_j - x_\ell \right|}{\lambda} \right), \tag{18}$$

where $\lambda > 0$ is the correlation length of the prior and γ is the prior marginal variance of each pixel. We set

$$\pi_{\mathrm{prior}}(\theta) \sim \mathcal{N}(0, \Gamma). \tag{19}$$

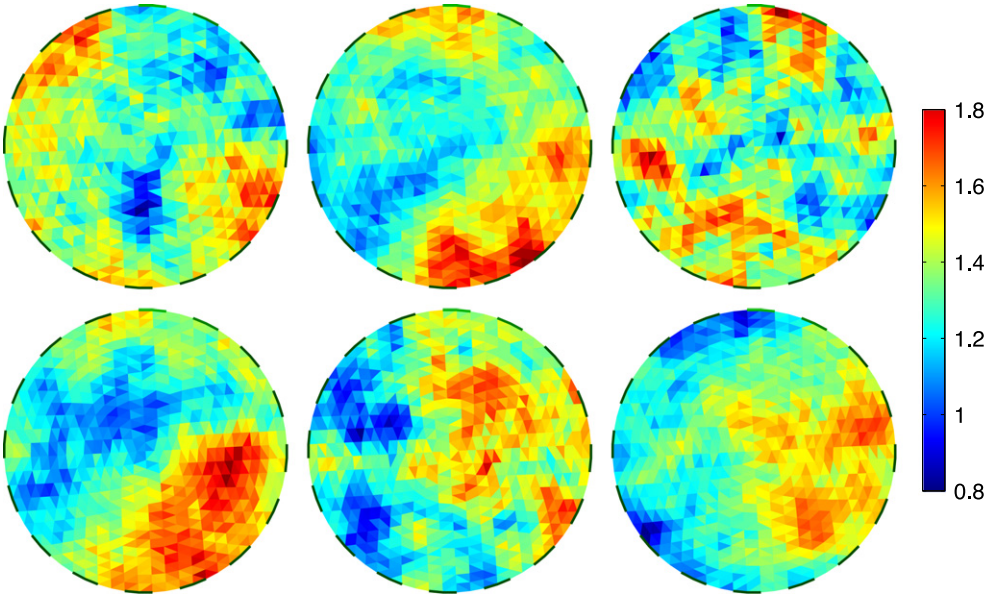

Figure 2. Six random realizations of the conductivity model (16) with the coefficient vector θ drawn from the prior distribution (19). The radius of the disc is 0.2 m, and the correlation length in (18) is $\lambda = 0.2$ m. The background conductivity is $\sigma_0 = 1.3\ \mathrm{S\,m^{-1}}$, and the prior variance is $\gamma = 0.05\ \mathrm{S^2\,m^{-2}}$, corresponding to a standard deviation of $\sqrt{\gamma} \approx 0.224\ \mathrm{S\,m^{-1}}$.

In principle, this prior allows the values of σ in (16) to become negative, and therefore we reject draws leading to negative conductivities. In practice, for γ sufficiently small, the rare negative values are not a problem. Figure 2 shows six conductivity distributions given by (16) with the vectors θ drawn from the prior (19).

3.4. Computed examples

The forward problem is implemented by generating the second order finite element model and the Jacobians with the public domain EIT research tool EIDORS ([25]). The inverse solver of step (iii) in algorithm 2 requires a non linear optimization algorithm, which we based on a standard Gauss–Newton iteration with backtracking.

The data is generated using a conductivity distribution defined in the fine grid, and shown in figure 3. The discrete, element-based representation of this conductivity is denoted by σ^*,

$$\sigma^*(x) = \sigma_0 + \sum_{j=1}^{N_e} \theta_j^* \chi_j^N(x),$$

where χ_j^N is the characteristic function of the jth element in the fine grid.

We will denote by $\mathcal{F}^N\colon \mathbb{R}^{N_e} \to \mathbb{R}^{L(L-1)}$ the forward map from the fine grid conductivity to voltages, while $f^N\colon \mathbb{R}^{n_e} \to \mathbb{R}^{L(L-1)}$ is the fine grid solver that first projects the conductivity from the coarse grid to the fine one. Formally, we can write $f^N = \mathcal{F}^N \circ P$, where $P\colon \mathbb{R}^{n_e} \to \mathbb{R}^{N_e}$ is the projection operator.

As pointed out above, the effect of the modeling error on the conductivity estimate is particularly noticeable when the exogenous noise such as imprecision in measurements is

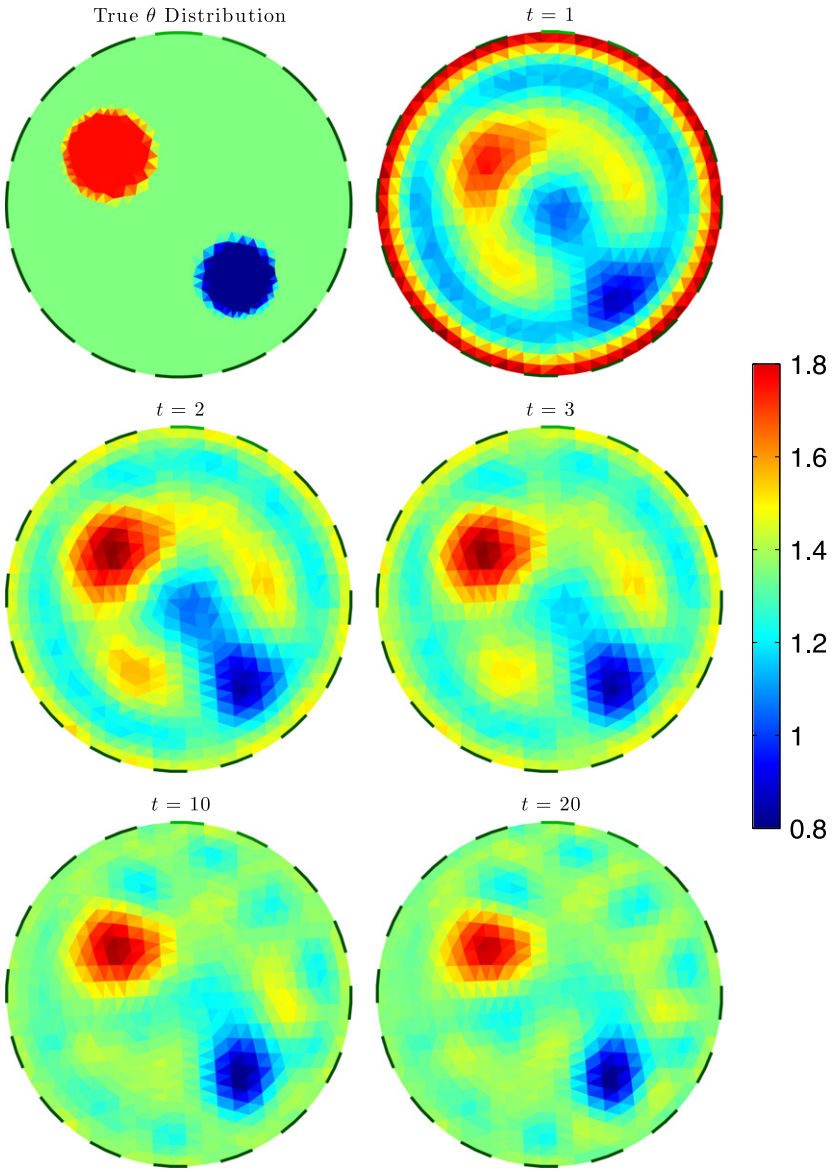

Figure 3. The conductivity distribution used to generate the data is shown in the top row, left column. The background conductivity is $\sigma_0 = 1.3\,\mathrm{S\,m^{-1}}$, and the maximum conductivity (red) and minimum conductivity (blue) are $1.6\,\mathrm{S\,m^{-1}}$ and $0.9\,\mathrm{S\,m^{-1}}$, respectively. The contact impedance value is fixed and assumed to be known, with $z = 0.003\,\Omega\mathrm{m}^2$. The figure shows the ensemble average conductivities with sample size $k = 20$ at different iteration rounds. The first estimate on the top left, computed with no error model, shows strong modeling error artifacts, in particular at the domain boundary. The ensemble mean stabilizes quickly, requiring less than ten iterations to remove the artifacts. Some of the background fluctuations are due to the small sample size.

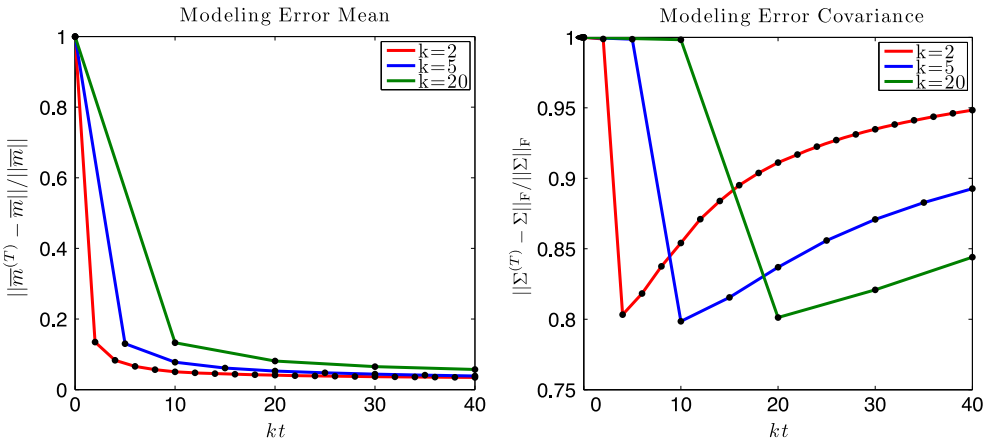

Figure 4. The convergence of the relative difference between the dynamic and prior-based modeling error mean (left) and covariance (right). The computation of the prior-based quantities is done with a fixed number $K = 2500$ of random draws from the prior. The horizontal axis indicates the total number of sample vectors in the cumulative dynamic mean and covariance. The plots show that the sample means obtained by the two methods are not very different, regardless of the sample size k, but the dynamic modeling error covariances are significantly smaller than the prior based estimate.

low, hence the modeling error dominates. Therefore, we assume a low noise level, and corrupt the synthetic data with zero mean Gaussian additive white noise with variance η^2. The standard deviation is set to be proportional to the norm of the difference of the computed noiseless data and the response that would correspond to a constant background conductivity,

$$\eta = \delta \parallel \mathcal{F}^N\left(\theta^*\right) - \mathcal{F}^N(0) \parallel,$$

where the constant of proportionality is chosen to be $\delta = 7 \times 10^{-3}$. The emphasis being on error modeling, this noise level is chosen low enough to make the modeling error the predominant part in the noise.

We start by computing a sequence of conductivity estimates with sample size $k = 20$. The sample averages after $t = 1, 2, 3, 10$ and 20 iterations are shown in figure 3. We see that in the first round ($t = 1$), when no discretization error model is included, the estimate shows a strong characteristic circular artifact which is particularly strong at the boundary where the coarse mesh is unable to accurately approximate the effect of the electrodes on the solutions. However, the algorithm quickly learns the statistics of the modeling error, and in less than ten iterations, the sample mean estimates become stable. The persistent fluctuations in the background conductivity are due to the small sample size.

It is of interest to see how the modeling error mean $\overline{m}^{(T)}$ and covariance $\Sigma^{(T)}$ evolve over the iterations, and in particular, how they compare to the sample mean $\overline{m}$ and covariance Σ based on draws from the prior density of θ. We compute approximations $\overline{m}$ and Σ as in step (ii) of algorithm 1 with $K = 2500$ random draws from the prior $\pi_{\text{prior}}(\theta)$ given by (19) and plot the relative differences $\parallel \overline{m}^{(T)} - \overline{m} \parallel / \parallel \overline{m} \parallel$ and $\parallel \Sigma - \Sigma^{(T)} \parallel_{\text{F}} / \parallel \Sigma \parallel_{\text{F}}$, where subindex indicates the Frobenius matrix norms. The results for different values of the sample size k in algorithm 2 are shown in figure 4. In the figures, the horizontal axis indicates the number of sample vectors required (kt), and can thus be thought of as a proxy for computation time.

The convergence of the sample mean $\overline{m}^{(T)}$ indicates that the estimates of the modeling error mean based on the prior draws, and on the approximate posterior sampling draws are not

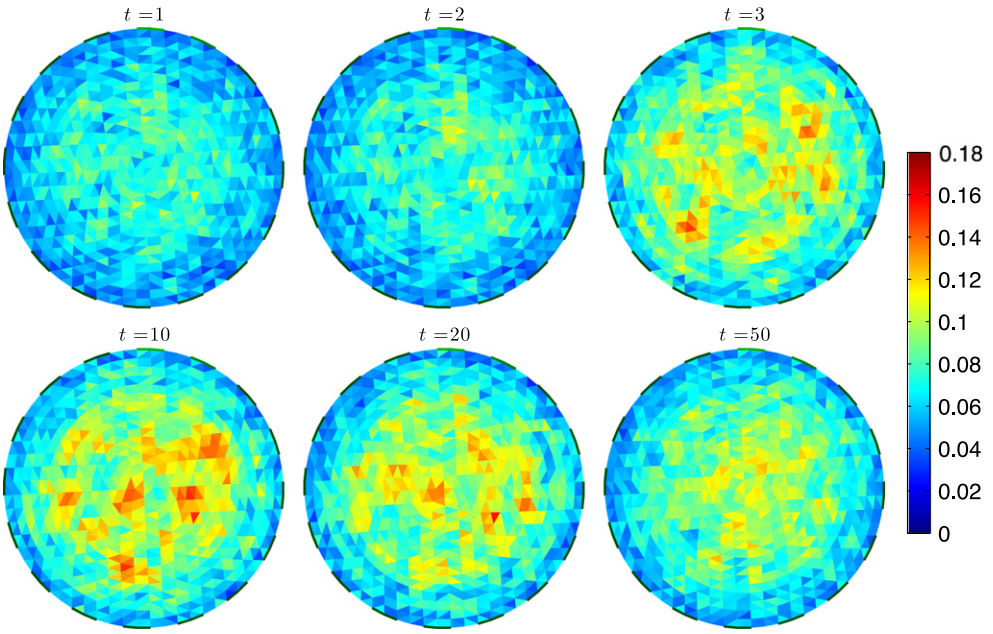

Figure 5. The plot of the square roots of the diagonal entries of the ensemble-based covariance matrix of the vector θ as a function of the iteration rounds. The sample size is $k = 20$.

significantly different. As t increases, the relative difference stabilizes to a level of about 5%. The convergence in number of iterations is insensitive to the sample size k and depends only on t.

The evolution of the modeling error covariance, on the other hand, is significantly different. After one iteration, the relative error is close to 100% regardless of the sample size k, and at the second step, the approximate covariance moves towards the prior-based covariance, only to deviate from it again as the iterations proceed. The behavior is similar for all sample sizes. A simple explanation of this phenomenon is that the first sample $\mathcal{S}_k^1$ consists of conductivity distributions that have no information about the modeling error, and the bootstrap data sample b_ℓ is drawn using the covariance of the very low level exogenous noise. In addition, the prior term has small impact on the randomized MAP estimates at the first iteration since the likelihood covariance has no modeling error correction and is therefore the dominant part of the posterior. Consequently, the set $\mathcal{S}_k^1$ contains almost identical solutions with a strong modeling error artifact, and the sample-based covariance approximately vanishes, thus $\Sigma - \Sigma^{(1)} \approx \Sigma$. At the second time step, the algorithm starts to learn about the modeling error, leading to a covariance more similar to Σ. As the iterations proceed, the samples $\mathcal{S}_k^t$ are approximations of random draws from the posterior density of θ, which has a significantly smaller variance than the prior. Consequently, the variance of the modeling error calculated from this sample is also significantly smaller than when the calculation is based on the wider prior density.

In figure 5, we visualize the time evolution of the square roots of the diagonal entries of the sample-based estimate of the covariance matrix of θ, which are approximations of the marginal standard deviations of the individual pixel values. The plots agree with the expectation that the uncertainty is larger in the center of the domain than near the boundary.

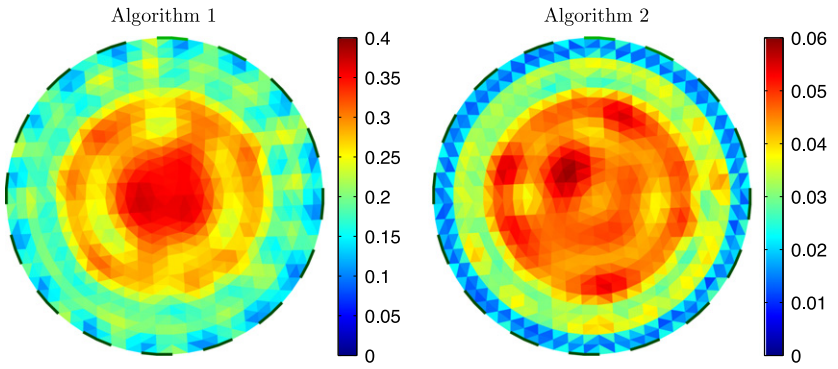

Figure 6. Estimated standard deviations of the pixels using the prior draws of θ to estimate the modeling error covariance (left), compared to the result by using the dynamic updating for the error covariance (right) after 20 time steps. The posterior covariance matrix of θ is approximated through sample averaging over an ensemble of 1000 realizations of θ computed with the randomized optimization sampling.

Due to the small sample size, $k = 20$, the images are noisy, but the variance reduction from $t = 10$ to $t = 50$ is visible.

The numerical experiments indicate that the modeling error variance with dynamic updating is significantly smaller than the variance estimated by random draws from the prior. This reduction of the likelihood variance leads to a reduced posterior variance for the parameter θ itself. Figure 6 shows an estimate of the posterior standard deviations of the pixels computed by the randomized optimization sampling using the prior-based estimate of the modeling error covariance on one hand, compared to the result with the dynamically updated estimate on the other. The plots show that the dynamic modeling error updating reduces the marginal standard deviations almost by an order of magnitude.

Our analysis indicates that satisfactory estimates for the modeling error statistics can be achieved with extremely small sample sizes ($k = 2$, $k = 5$), leading to a fast algorithm with no need of off-line computations. The very small sample size, however, may result in noisier sample mean estimates. In figure 7, the sample means after $T = 20$ iterations with different sample sizes are shown, together with the MAP estimate calculated by algorithm 1 using the off-line modeling error statistics estimate with $K = 2500$ random draws from the prior. In the latter estimate, the background is less noisy than in the dynamic sample mean estimates, however the inclusions are slightly more diffuse. We point out that while the modeling error estimate in algorithm 1 is sample-based, the conductivity estimate is a single MAP estimate, and in order to estimate of the posterior covariance, a separate sampling algorithm needs to be implemented.

4. Conclusions

This article describes a novel sampling-based algorithm for inverse problems with discretization-based model discrepancies, based on the idea of dynamically updating in an alternating fashion the parameter sample and the modeling error statistics. The viability of this algorithm is demonstrated with an application to the inverse problem of EIT. Unlike the prior-sampling based discretization error modeling proposed in the literature, our algorithm does not require off-line computations, and the adequacy of the sample size can be assessed as the

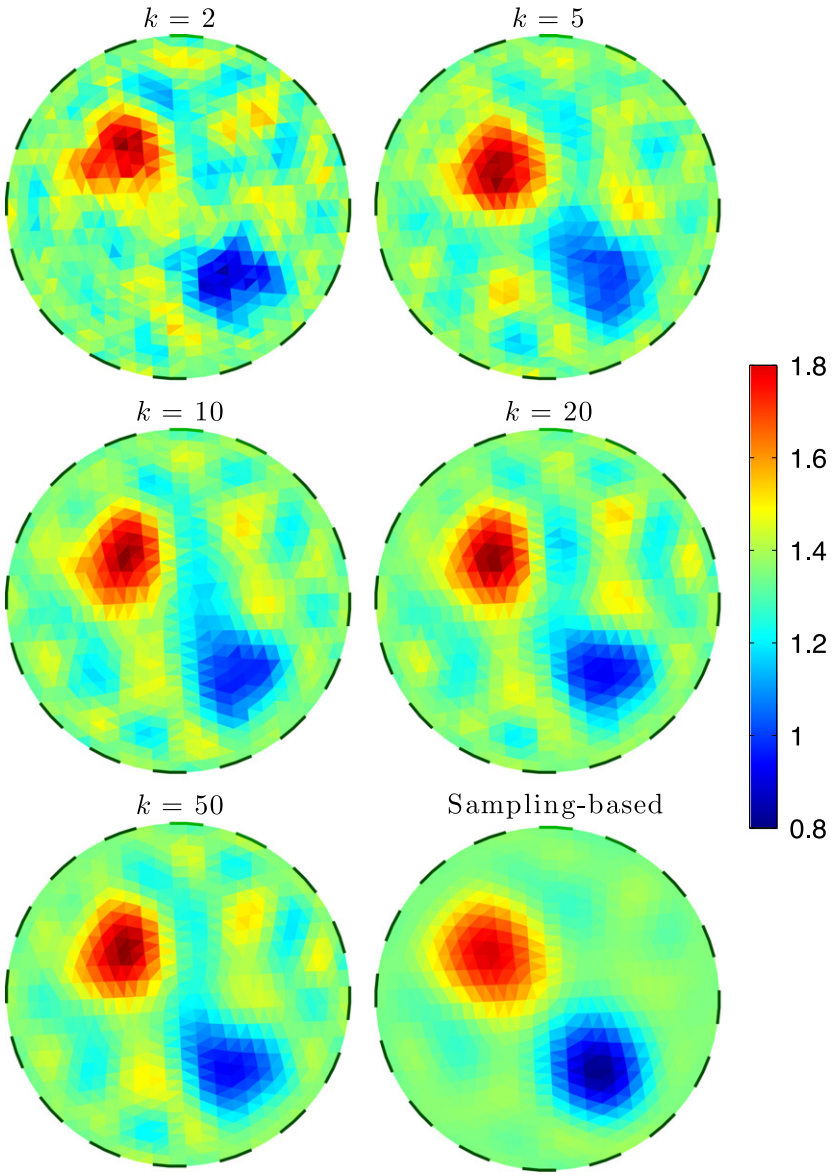

Figure 7. Final sample mean estimates after $T = 20$ iteration rounds with different sample sizes, $k = 2$, 5, 10, 20 and 50. For comparison, the conductivity estimate at the bottom of the right column is computed by estimating the modeling error statistics with an off-line random draws from the prior $\pi_{\text{prior}}(\theta)$. The sample size for that computation is $K = 2500$.

iterations proceed. Moreover, the speed of the algorithm can be adjusted by controlling the sample size, but the trade-off for decreasing the computation time is that the sample-based mean estimate, as well as the estimate of the posterior covariance may become more noisy. Computed examples demonstrate that even with extremely low sample size, less than ten, the sample-based estimates are of reasonably good quality. Preliminary numerical tests indicate

that the dynamical updating of the modeling error statistics works also if the target conductivity is not kept fixed, indicating that the algorithm can be adapted to time-varying inverse problems, in which the forward model needs not be stationary. This observation allows the extension of the proposed error estimate methodology to genuinely dynamic models.

Acknowledgments

The work was partly supported by the NSF (DMS grant 1312424 to Erkki Somersalo) and by the Simons Foundation (grant # 246665 to Daniela Calvetti).

References

[1] Arridge S R, Kaipio J P, Kolehmainen V, Schweiger M, Somersalo E, Tarvainen T and Vauhkonen M 2006 Approximation errors and model reduction with an application in optical diffusion tomography *Inverse Problems* **22** 175–95
[2] Brynjarsdóttir J and O'Hagan A 2014 Learning about physical parameters: the importance of model discrepancy *Inverse Problems* **30** 114007
[3] Calvetti D and Somersalo E 2005 Statistical compensation of boundary clutter in image deblurring *Inverse Problems* **21** 1697–714
[4] Calvetti D, McGivney D and Somersalo E 2012 Left and right preconditioning for electrical impedance tomography with structural information *Inverse Problems* **28** 055015
[5] Cheney M, Isaacson D and Newell J C 1999 Electrical impedance tomography *SIAM Rev.* **41** 85–101
[6] Cui T, Fox C and O'Sullivan M J 2011 Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm *Water Resour. Res.* **47** W10521
[7] Efron B 1979 Bootstrap methods: another look at the jackknife *Ann. Stat.* **7** 1–26
[8] Efron B and Tibshirani R 1993 *An Introduction to the Bootstrap* (London: Chapman and Hall)
[9] Evensen G 1994 Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte–Carlo methods to forecast error statistics *J. Geophys. Res.* **99** 10143–62
[10] Heino J and Somersalo E 2004 A modelling error approach for the estimation of optical absorption in the presence of anisotropies *Phys. Med. Biol.* **49** 4785–98
[11] Heino J, Somersalo E and Kaipio J 2005 Statistical compensation of geometric mismodeling in optical tomography *Opt. Express* **13** 296–308
[12] Higdon D, Gattiker J, Williams B and Rightley M 2008 Computer model calibration using high-dimensional output *J. Amer. Stat. Assoc.* **103** 103–482
[13] Iglesias M A, Law K J H and Stuart A M 2013 Ensemble Kalman methods for inverse problems *Inverse Problems* **29** 045001
[14] Kaipio J P and Somersalo E 1999 Nonstationary inverse problems and state estimation *J. Inv. Ill-Posed Probl.* **7** 273–82
[15] Kaipio J P and Somersalo E 2004 *Statistical and Computational Inverse Problems* (New York: Springer)
[16] Kaipio J P and Somersalo E 2007 Statistical inverse problems: discretization, model reduction and inverse crimes *J. Comp. Appl. Math.* **198** 493–504
[17] Kennedy M C and O'Hagan A 2011 Bayesian calibration of computer models *J. R. Stat. Soc. Ser.* B **63.3** 425–64
[18] Lipponen A, Seppänen A and Kaipio J P 2011 Nonstationary approximation error approach to imaging of three-dimensional pipe flow: experimental evaluation *Meas. Sci. Technol.* **22** 104013
[19] Nissinen A, Kolehmainen V and Kaipio J P 2011 Compensation of modelling errors due to unknown domain boundary in electrical impedance tomography *IEEE Trans. Med. Imaging* **30** 231–42

[20] Nissinen A, Kolehmainen V and Kaipio J P 2011 Reconstruction of domain boundary and conductivity in electrical impedance tomography using the approximation error approach *Int. J. Uncertain. Quantification* **1** 203–22
[21] Oliver D S, Reynolds A C and Liu N 2008 *Inverse Theory for Petroleum Reservoir Characterization and History Matching* (Cambridge: Cambridge University Press)
[22] Oliver D S and Chen Y 2011 Recent progress on reservoir history matching: a review *Comput. Geosci.* **15** 185–221
[23] Somersalo E, Isaacson D and Cheney M 1992 Existence and uniqueness for electrode models for electric current computed tomography *SIAM J. Appl. Math.* **52** 1023–40
[24] Tarvainen T, Kolehmainen V, Pulkkinen A, Vauhkonen M, Schweiger M, Arridge S R and Kaipio J P 2010 An approximation error approach for compensating for modelling errors between the radiative transfer equation and the diffusion approximation in diffuse optical tomography *Inverse Problems* **26** 015005
[25] Vauhkonen M, Lionheart W R, Heikkinen L M, Vauhkonen P J and Kaipio J P 2001 A MATLAB package for the EIDORS project to reconstruct two-dimensional EIT images *Physiol. Meas.* **22** 107