

Parameterizations for Ensemble Kalman Inversion

Neil K. Chada^{*}, Marco A. Iglesias^{**}, Lassi Roininen[†]
and Andrew M. Stuart[‡]

^{*} Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK.

^{**} School of Mathematical Sciences, University of Nottingham,
Nottingham Park, NG7 2QL, UK.

[†] Department of Mathematics, Imperial College London, London, SW7 2RD, UK.

[‡] Computing and Mathematics Sciences, California Institute of Technology,
Pasadena, CA 91125, USA.

E-mail:

n.chada@warwick.ac.uk

marco.iglesias@nottingham.ac.uk

l.roininen@imperial.ac.uk

astuart@caltech.edu

Abstract. The use of ensemble methods to solve inverse problems is attractive because it is a derivative-free methodology which is also well-adapted to parallelization. In its basic iterative form the method produces an ensemble of solutions which lie in the linear span of the initial ensemble. Choice of the parameterization of the unknown field is thus a key component of the success of the method. We demonstrate how both geometric ideas and hierarchical ideas can be used to design effective parameterizations for a number of applied inverse problems arising in electrical impedance tomography, groundwater flow and source inversion. In particular we show how geometric ideas, including the level set method, can be used to reconstruct piecewise continuous fields, and we show how hierarchical methods can be used to learn key parameters in continuous fields, such as length-scales, resulting in improved reconstructions. Geometric and hierarchical ideas are combined in the level set method to find piecewise constant reconstructions with interfaces of unknown topology.

PACS numbers: 62M20, 49N45, G5L09.

Keywords: Ensemble Kalman inversion, hierarchical inversion, centered and non-centered parameterizations, discontinuous reconstructions, level set methods.

Submitted to: *Inverse Problems*

1. Introduction

1.1. Content

Consider finding u from y where

$$y = \mathcal{G}(u) + \eta, \quad (1)$$

\mathcal{G} is a forward map taking the unknown parameter u into the data space, and η represents noise. Ensemble Kalman inversion is an attractive technique that has shown considerable success in the solution of such problems. Whilst it is derived from the application of Kalman-like thinking, with means and covariances computed from an empirical ensemble, it essentially acts as a black-box derivative-free optimizer which requires only evaluation of the forward map $\mathcal{G}(\cdot)$; in practice it can often return good solutions to inverse problems with relatively few forward map evaluations. However the choice of parameterization of the unknown is key to the success of the method. In this paper we will demonstrate how carefully thought out parameterizations can have substantial impact in the quality of the reconstruction.

Although our viewpoint in this paper is to consider Ensemble Kalman inversion as an optimization method, and evaluate it from this perspective, there is considerable insight to be gained from the perspective of Bayesian inversion; this is despite the fact that the algorithm does not, in general, recover or sample the true Bayesian posterior distribution of the inverse problem. Algorithms that can, with controllable error, approximately sample from the true posterior distribution are commonly referred to as fully Bayesian, with examples including Markov Chain Monte Carlo and sequential Monte Carlo. Ensemble Kalman inversion is not fully Bayesian but the link to Bayesian inversion remains important as we now explain. There is considerable literature available about methods to improve fully Bayesian approaches to the inverse problem through for example, geometric and hierarchical parameterizations of the unknown. The purpose of this paper is to demonstrate how these ideas from Bayesian inversion may be used with some success to improve the capability of ensemble Kalman methods considered as optimizers. In view of the relatively low computational cost of the ensemble methods, in comparison with fully Bayesian inversion, this cross-fertilization of ideas has the potential to be quite fruitful.

1.2. Literature Review

The Kalman filter (KF) [30] was developed to sequentially update the probability distribution on states of partially observed linear Gaussian systems, and subsequently generalized to nonlinear problems in the form of the extended Kalman filter. However for high-dimensional systems the size of covariances makes use of these methods prohibitive. In 1994 Evensen [19, 20] proposed a Monte-Carlo based nonlinear Kalman filter which tackled this issue by using an ensemble of particles to represent the covariances and mean, resulting in what is now known as the ensemble Kalman filter (EnKF). A major

success story for the EnKF has been in weather prediction models [2, 23], but it has also been deployed in numerous applications domains, including the reservoir engineering community [1] and in oceanography [21]. Variants on the idea include the randomized maximum likelihood (RML) method [40], and algorithms such as the ensemble square-root Kalman filter [53].

In this paper we are primarily interested in use of ensemble Kalman methods to study inverse problems for parameter estimation, an approach pioneered for oil industry applications where the inverse problem is known as history matching [32, 40]; the paper [18] contains an insightful analysis of the methodology in the large ensemble limit. In this application domain such inversion methods are sometimes referred to as ensemble Kalman smoothers, although the nomenclature is not uniform. We will simply refer to *ensemble Kalman inversion* (EKI). The methodology is formulated quite generally in [26], independently of oil industry applications; and in [25] it is shown that the method performs well as an optimizer but does not capture true posterior uncertainty, in the context of oil industry applications.

The ideas introduced in this paper concerning parameterization are independent of the particular implementation of the EKI method used; in all our numerical experiments we will use the form of iterative regularization proposed by Iglesias [24]. The general philosophy behind the method is that, as the algorithm is iterated, the solution to the inverse problem should approach the truth in the small noise limit, and hence that any regularization introduced should diminish in influence in this limit. However, because the convergence theory for ensemble Kalman inversion is in its infancy, the choice of iterative regularization is made by analogy with classical iterative methods that have been used for inverse problems [5, 6, 22], with the ensemble method using empirical covariances in place of derivatives or adjoints of the forward solver. The resulting iterative method is an ensemble version of the Levenberg-Marquardt algorithm with the inclusion of regularization as in [22].

In the form of EKI that we use the linear span of the initial ensemble is preserved by the iteration [26, 32]. The initial ensemble thus encodes prior information about the solution of the problem. This means that choice of the parameterization of the method, as well as the choice initial subspace, is key to its performance. Based on experience with (Bayesian) statistical modelling we will introduce geometric and hierarchical priors that address the issue of making good parameterizations, and we will draw from those priors to create the initial ensemble. Hierarchical models have been extensively studied in the fields of computational statistics and machine learning [36, 41, 50]. Their use in the context of Monte Carlo Markov chain (MCMC) methods for Bayesian inverse problems is overviewed in [42]; see [16, 46] for application oriented work. One important outcome of research in this area is that learning parameters such as length-scale, amplitude and regularity within Gaussian random field priors (such as Whittle-Matérn) can be of significant value [4, 38, 45]. In a series of recent papers this hierarchical modelling was extended to allow for length-scale which is itself spatially varying [34, 45]. The development of hierarchical methods within EKI, rather than fully Bayesian MCMC,

has been limited to date, with the primary contribution being the work [17] where the methodology was based on building large ensembles from multiple Gaussians assigned different weights. However this work requires that the correct hierarchical parameter is in the ensemble if it is to be successful. We also note that there is some work in hierarchical EnKF within the context of state estimation; see [51] and the references therein.

In addition to hierarchical approaches we will also study geometric parameterizations. These can be of use when the geometric object has known form, such as faults, channels and inclusions, or when it is of unknown topology the level set method may be used [47]. We will build on recent Bayesian implementation of these ideas; see [27, 28] and references therein.

1.3. Contribution of This Work

Our main contribution is to establish the importance of novel parameterizations which have the potential to substantially improve the performance of EKI. Although our perspective on EKI is one of optimization, the methods we introduce are all based on taking established and emerging methods from Bayesian statistics and developing them in the context of the ensemble methods. The connection to Bayesian statistics is exploited to provide insights about how to make these methods efficient. The resulting methods are illustrated by means of examples arising in both electrical impedance tomography, groundwater flow and source inversion. The contributions are:

- We develop hierarchical approaches for EKI, based on solving for the unknown function and unknown scalars which parameterize the prior.
- We generalize these hierarchical approaches to EKI to include unknown fields which parameterize the prior, rather than scalars.
- We demonstrate the key role of choosing non-centered variables when implementing hierarchical methods.
- We show the potential for geometric hierarchical priors, including the level set parameterization, for piecewise continuous reconstructions.

1.4. Organization

The layout of the paper is as follows, in section 2 we discuss different approaches to parameterizing inverse problems. We begin by conveying the main ideas in subsection 2.1 in abstract. In subsection 2.2 we describe these ideas more concretely. In section 3 we describe the hierarchical version of iterative EKI as used in this paper, and section 4 describes the model problems that we use to illustrate the power of the proposed parameterizations. Numerical results are presented in section 5, whilst in section 6 we make some concluding remarks.

1.5. Notation

Throughout the paper we make use of common notation for Hilbert space norms and inner products, $\|\cdot\|$, $\langle\cdot,\cdot\rangle$. We will assume that \mathcal{X} and \mathcal{Y} are two separable Hilbert spaces which are linked through the forward operator $\mathcal{G} : \mathcal{X} \rightarrow \mathcal{Y}$. This nonlinear operator can be thought of as mapping from the space of unknown parameters \mathcal{X} to the observation space \mathcal{Y} . Our additive noise for the inverse problems will be denoted by $\eta \sim N(0, \Gamma)$ where $\Gamma : \mathcal{Y} \rightarrow \mathcal{Y}$ is a self-adjoint positive operator. For any such operator we define $\langle\cdot,\cdot\rangle_\Gamma = \langle\Gamma^{-1/2}\cdot,\Gamma^{-1/2}\cdot\rangle$ and $\|\cdot\|_\Gamma = \|\Gamma^{-1/2}\cdot\|$, and for finite dimensions $|\cdot|_\Gamma = |\Gamma^{-1/2}\cdot|$ with $|\cdot|$ the Euclidean norm. If the Gaussian measure associated to η is supported on \mathcal{Y} then we will require Γ to be trace-class; however we will also consider white noise whose support is on a larger space than \mathcal{Y} and for which the trace-class condition fails in the infinite dimensional setting.

2. Inverse Problem

2.1. Main Idea

2.1.1. Non-Hierarchical Inverse Problem We are interested in the recovery of $u \in \mathcal{X}$ from measurements of $y \in \mathcal{Y}$ given by equation (1) in which, recall, η is additive Gaussian noise. In the Bayesian approach to inverse problems we treat each quantity within (1) as a random variable. Via an application of Bayes' Theorem ‡ [13] we can characterize the conditional distribution of $u|y$ via

$$\mathbb{P}(u|y) \propto \mathbb{P}(y|u) \times \mathbb{P}(u), \quad (2)$$

where $\mathbb{P}(u)$ is the prior distribution, $\mathbb{P}(u|y)$ is the posterior distribution and $\mathbb{P}(y|u)$ is the likelihood. Although we view EKI as an optimizer in this paper, the Bayesian formulation of (1) is important because we derive the initial ensemble from the prior distribution $\mathbb{P}(u)$. From an optimization viewpoint, our goal is to make the following least squares objective function small:

$$\Phi(u; y) = \frac{1}{2}|y - \mathcal{G}(u)|_\Gamma^2. \quad (3)$$

This (upto an irrelevant additive constant) is the negative log likelihood since it is assumed that η is a mean-zero Gaussian with covariance Γ .

2.1.2. Centered Hierarchical Inverse Problem In many applications it can be advantageous to add additional unknowns θ to the inversion process. In particular these may enter through the prior, as in hierarchical methods, [42], and we will refer to such parameters as hyperparameters. The inverse problem is then the recovery of

‡ We write all instances of Bayes' Theorem in finite dimensions for simplicity; extension to Bayes' Theorem for functions is straightforward but not central to this paper and so we avoid the extra notation that would be needed for this.

(u, θ) from measurements of y given again by (1). The additional parameterization of the prior results in Bayes' Theorem in the form

$$\mathbb{P}(u, \theta|y) \propto \mathbb{P}(y|u) \times \mathbb{P}(u, \theta). \quad (4)$$

Prior samples, used to initialize the ensemble smoother, will then be of the pair (u, θ) . What we will term *centered* hierarchical methods (a terminology we discuss in Section 3) typically involve factorization of the prior in the form $\mathbb{P}(u, \theta) = \mathbb{P}(u|\theta)\mathbb{P}(\theta)$. From an optimization point of view our goal is again to make the objective function (3) small, but now using hierarchical parameterization to construct the initial ensemble. §

2.1.3. Non-Centered Hierarchical Inverse Problem Another variant of the inverse problem that is particularly relevant for hierarchical methods, in which θ enters only the prior, is *non-centered* reparameterization (a further terminology discussed in Section 3). We introduce the transformation $T : (\xi, \theta) \rightarrow u$ and note that (1) then becomes

$$y = \mathcal{G}(T(\xi, \theta)) + \eta, \quad (5)$$

and Bayes' Theorem then reads

$$\mathbb{P}(\xi, \theta|y) \propto \mathbb{P}(y|\xi, \theta) \times \mathbb{P}(\xi, \theta). \quad (6)$$

Prior samples, again used to initialize the ensemble smoother, will then be of the pair (ξ, θ) . Typically the change of variables from u to ξ is introduced so that ξ and θ are independent under the prior: $\mathbb{P}(\xi, \theta) = \mathbb{P}(\xi)\mathbb{P}(\theta)$. As a result the inverse problem (5) is different to that appearing in (1), in terms of both the prior and the likelihood. This non-centered approach is equivalent to the ancillary augmentation technique discussed in [41] which discusses the decoupling of u and θ through the variable ξ . From an optimization viewpoint, our goal is to make the following least squares objective function small:

$$\Phi(\xi, \theta; y) = \frac{1}{2} |y - \mathcal{G}(T(\xi, \theta))|_{\Gamma}^2. \quad (7)$$

In this paper we consider the application of EKI for solution of the inverse problems (1), non-hierarchically and centered-hierarchically, and (5) non-centered hierarchically. Although this method has a statistical derivation, the work in [26, 35] demonstrates that the method may be thought of as a derivative-free optimizer that approximates the least squares problem (3) and (7) rather than sampling from the relevant Bayesian posterior distribution. We will show that the use of iterative EKI, as proposed by Iglesias in [24], can effectively solve a wide range of challenging inversion problems, if judiciously parameterized.

§ We note that hyperparameters θ may also enter the likelihood as well as the prior if the state variable is re-scaled in a hyperparameter-dependent fashion, as happens in the version of the Bayesian level set method advocated in [16].

2.2. Details of Parameterizations

In this section we describe in detail several classes of parameterizations that we use in this paper. The first and second are geometric parameterizations, ideal for piecewise continuous reconstructions with unknown interfaces. The third and fourth are hierarchical methods which introduce an unknown length-scale, and regularity parameter, into the inversion. For the two geometric problems we initially formulate in terms of trying to find a function $w : D \mapsto \mathbb{R}$, D a subset of \mathbb{R}^d , and then reparameterize w . For the hierarchical problems we initially formulate in terms of trying to find a function $u : D \mapsto \mathbb{R}$, and then append parameters θ and also rewrite in terms of $(\xi, \theta) \mapsto u$.

2.2.1. Geometric Approach – Finite Dimensional Parameterization In many problems of interest the unknown function w has discontinuities, determination of which forms part of the solution of the inverse problem. To tackle such problems it may be useful to write w in the form

$$w(x) = \sum_{i=1}^n u_i(x) \chi_{D_i}(x). \quad (8)$$

Here the union of the disjoint sets D_i is the whole domain D . If we assume that the configuration of the D_i is determined by a finite set of scalars θ and let u denote the union of the functions u_i and the parameters θ then we may rewrite the inverse problem in the form (1). The case where the number of subdomains n is unknown would be an interesting and useful extension of this work; but we do not consider it here.

2.2.2. Geometric Approach – Infinite Dimensional Parameterization If the interface boundary is not readily described by a finite number of parameters we may use the level set idea. For example if the field w takes two known values w^\pm with unknown interfaces between them we may write

$$w(x) = w^+ \mathbb{I}_{u>0}(x) + w^- \mathbb{I}_{u<0}(x), \quad (9)$$

and formulate the inverse problem in the form (1) for u . This idea may be generalized to functions which take an arbitrary number of constant values, through the introduction of level sets other than $u = 0$, or through vector level sets functions u .

2.2.3. Scalar-valued Hierarchical Parameterizations To illustrate ideas we will concentrate on Gaussian priors of Whittle-Matérn type. These are characterized by a covariance function of the form

$$c(x, y) = \sigma^2 \frac{2^{1-\alpha+d/2}}{\Gamma(\alpha - d/2)} \left(\frac{|x - y|}{\ell} \right)^{\alpha-d/2} K_{\alpha-d/2} \left(\frac{|x - y|}{\ell} \right), \quad x, y \in \mathbb{R}^d, \quad (10)$$

where K is a modified Bessel function of the second kind, $\sigma^2 > 0$ is the variance and $\Gamma(\cdot)$ is a Gamma function. We will always ensure that $\ell > 0$ and $\alpha > d/2$ so that

draws from the Gaussian are well-defined and continuous. On the unbounded domain \mathbb{R}^d , samples from this process may be generated by solving the stochastic PDE

$$(I - \ell^2 \Delta)^{\frac{\alpha}{2}} u = \ell^{d/2} \sqrt{\beta} \xi, \quad (11)$$

where $\xi \in H^{-s}(D)$, $s > \frac{d}{2}$, is a Gaussian white noise, i.e. $\xi \sim N(0, I)$, and

$$\beta = \sigma^2 \frac{2^d \pi^{d/2} \Gamma(\alpha)}{\Gamma(\alpha - \frac{d}{2})}.$$

In this paper we will work with the scalar hierarchical parameters α and $\tau = \ell^{-1}$. Putting $\ell = \tau^{-1}$ into (11) gives the stochastic PDE

$$\mathcal{C}_{\alpha, \tau}^{-\frac{1}{2}} u = \xi, \quad (12)$$

where the covariance operator $\mathcal{C}_{\alpha, \tau}$ has the form $\mathcal{C}_{\alpha, \tau} = \tau^{2\alpha-d} \beta (\tau^2 I - \Delta)^{-\alpha}$. Throughout this paper we choose $\beta = \tau^{d-2\alpha}$ so that

$$\mathcal{C}_{\alpha, \tau} = (\tau^2 I - \Delta)^{-\alpha}, \quad (13)$$

and we equip the operator Δ with Dirichlet boundary conditions on D . These choices simplify the exposition, but are not an integral part of the methodology; different choices could be made.

We have thus formulated the inverse problem in the form of the centered hierarchical inverse problem (1) for (u, θ) with $\theta = (\alpha, \tau)$. The parameter θ enters only through the prior as in this particular case it does not appear in the likelihood. In this paper we will place uniform priors on α and τ , which will be specified in subsection 5.1. We may also work with the variables (ξ, θ) noting that (12) defines a map $T : (\xi, \theta) \mapsto u$ and we have formulated the inverse problem in the form (5), the non-centered hierarchical form. In Figures 1 and 2 we display random samples from (13) with imposed Dirichlet boundary conditions and varying values of the inverse length scale τ and the regularity α . These samples are constructed in the domain $D = [0, 1]^2$.

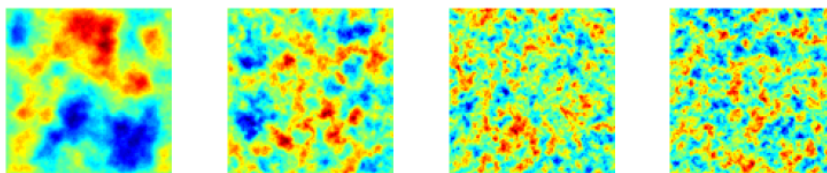


Figure 1. Modified inverse length scale for $\tau = 10, 25, 50$ and 100 . Here $\alpha = 1.6$.

2.2.4. Function-Valued Hierarchical Parameterization In order to represent non-stationary features it is of interest to allow hierarchical parameters to themselves vary in space. To this end we will also seek to generalize (11) and work with the form

$$(I - \ell(x; v)^2 \Delta)^{\frac{\alpha}{2}} u = \ell(x; v)^{\frac{d}{2}} \xi. \quad (14)$$

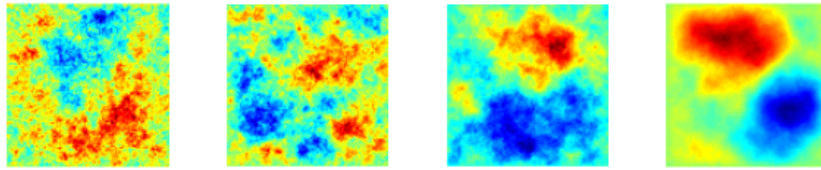


Figure 2. Modified regularity for $\alpha = 1.1, 1.3, 1.5$ and 1.9 . Here $\tau = 15$.

(We have set $\beta = 1$ for simplicity). In order to ensure that the length-scale is positive we will write it in the form

$$\ell(x; v) = g(v(x)), \quad (15)$$

for some positive monotonic increasing function $g(\cdot)$. We thus have a formulation as a centered hierarchical inverse problem in the form (1) noting that hyperparameter $\theta = v$ is here a function and enters only through the prior. We may also formulate inversion in terms of the variables (ξ, v) giving the inverse problem in the form (5) with $\theta = v$. We will consider two forms of prior on v . The first is based on a Gaussian random field with Whittle-Matérn covariance function (10) and we then choose $g(v) = \exp(v)$. The second, which will apply only in one dimension, is to consider a one-dimensional Cauchy process, as in [34]. In particular we will construct $\ell(x; v)$ by employing a one-dimensional Cauchy process $v(x)$, which is an α -stable Lévy motion with $\alpha = 1$ with Cauchy increments on the interval δ given by the density function f , and positivity-inducing function g , where

$$f(x) = \frac{\delta}{\pi(\delta^2 + x^2)}, \quad g(s) = \frac{a}{b + c|s|} + d, \quad (16)$$

such that $a, b, c, d > 0$ are constants. || Samples from these two priors on v , and hence ℓ , are shown in Figures 3 and 4.

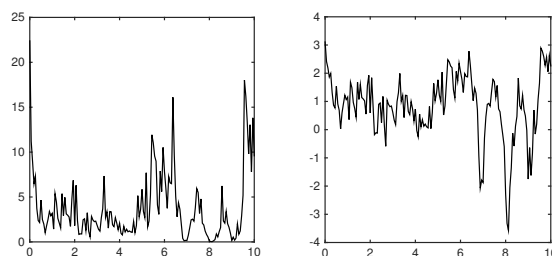


Figure 3. Gaussian random field. Left: Length-scale realization $\ell(x)$. Right: Realization of $v(x)$.

|| We note here that α has a different meaning from the parameter α used in the covariance function of a Gaussian prior in, for example, (13). We abuse notation in this way because the parameter α is widely used in the literature in both contexts.

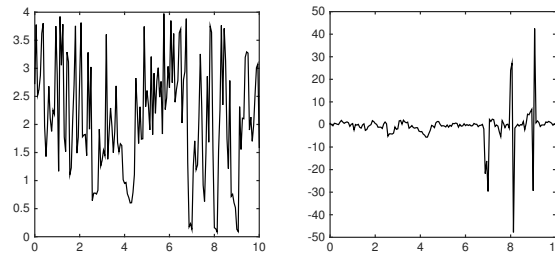


Figure 4. Cauchy random field. Left: Length-scale realization $\ell(x)$. Right: Realization of $v(x)$.

3. Iterative Ensemble Kalman Inversion

In this section we describe iterative EKI as implemented in this paper. We outline it first for inverse problems as parameterized in equations (1) and then discuss generalizations to centered hierarchical inversion and the non-centered hierarchical inversion (5). We briefly mention the continuous time limits of the methods as these provide insight into how EKI works, and the effect of re-parameterizing; the study of continuous limits from EKI was introduced in [48] and further details concerning their application to the problems considered here may be found in [12].

3.1. Formulation for (1)

The form of iterative EKI that we use is that employed in [24]. When applied to the inverse problem (1) it takes the following form, in which the subscript n denotes the iteration step, and the superscript (j) the ensemble member:

$$u_{n+1}^{(j)} = u_n^{(j)} + C_n^{uw} (C_n^{ww} + \Upsilon_n \Gamma)^{-1} (y - \mathcal{G}(u_n^{(j)})). \quad (17)$$

The empirical covariances C_n^{uw}, C_n^{ww} are given by

$$C_n^{uw} = \frac{1}{J-1} \sum_{j=1}^J (u_n^{(j)} - \bar{u}_n) \otimes (\mathcal{G}(u_n^{(j)}) - \bar{\mathcal{G}}_n) \quad (18)$$

$$C_n^{ww} = \frac{1}{J-1} \sum_{j=1}^J (\mathcal{G}(u_n^{(j)}) - \bar{\mathcal{G}}_n) \otimes (\mathcal{G}(u_n^{(j)}) - \bar{\mathcal{G}}_n). \quad (19)$$

Here \bar{u}_n denotes the average of $u_n^{(j)}$ over all ensemble members and $\bar{\mathcal{G}}_n$ denotes the average of $\mathcal{G}(u_n^{(j)})$ over all ensemble members. The parameter Υ_n is chosen to ensure that

$$\|y - \bar{\mathcal{G}}_n\|_{\Gamma} \leq \zeta \eta, \quad (20)$$

a form of discrepancy principle which avoids over-fitting.

If we define

$$d_n^{(j,m)} = \langle (C_n^{uw} + \Upsilon_n \Gamma)^{-1} (\mathcal{G}(u_n^{(j)}) - y), \mathcal{G}(u_n^{(m)}) - \bar{\mathcal{G}}_n \rangle,$$

Algorithm 1 Hierarchical Iterative Kalman Method (Centered Version).

Let $\{u_0^{(j)}, \theta_0^{(j)}\}_{j=1}^J \subset \mathcal{X}$ be the initial ensemble with J elements.
Further let $\rho \in (0, 1)$ with $\zeta > \frac{1}{\rho}$ and $\theta = (\alpha, \tau)$.

Generate $\{u_0^{(j)}, \theta_0^{(j)}\}$ i.i.d. from the prior $\mathbb{P}(u, \theta)$, with synthetic data
 $y_{n+1}^{(j)} = y + \eta_{n+1}^{(j)}$, $\eta^{(j)} \sim N(0, \Gamma)$ i.i.d.

Then for $n = 1, \dots$

1. **Prediction step:** Evaluate the forward map $w_n^{(j)} = \mathcal{G}(u_n^{(j)})$,

and define $\bar{w}_n = \frac{1}{J} \sum_{j=1}^J w_n^{(j)}$.

2. **Discrepancy principle:** If $\|\Gamma^{-1/2}(y - \bar{w}_n)\|_{\mathcal{Y}} \leq \zeta\eta$, stop!

Output $\bar{u}_n = \frac{1}{J} \sum_{j=1}^J u_n^{(j)}$ and $\bar{\theta}_n = \frac{1}{J} \sum_{j=1}^J \theta_n^{(j)}$.

3. **Analysis step:** Define sample covariances:

$$\begin{aligned} C_n^{uw} &= \frac{1}{J-1} \sum_{j=1}^J (u^{(j)} - \bar{u}) \otimes (\mathcal{G}(u^{(j)}) - \bar{\mathcal{G}}), \\ C_n^{\theta w} &= \frac{1}{J-1} \sum_{j=1}^J (\theta^{(j)} - \bar{\theta}) \otimes (\mathcal{G}(u^{(j)}) - \bar{\mathcal{G}}), \\ C_n^{ww} &= \frac{1}{J-1} \sum_{j=1}^J (\mathcal{G}(u^{(j)}) - \bar{\mathcal{G}}) \otimes (\mathcal{G}(u^{(j)}) - \bar{\mathcal{G}}). \end{aligned}$$

Update each ensemble member as follows

$$\begin{aligned} u_{n+1}^{(j)} &= u_n^{(j)} + C_n^{uw} (C_n^{ww} + \Upsilon_n \Gamma)^{-1} (y_{n+1}^{(j)} - \mathcal{G}(u_n^{(j)})), \\ \theta_{n+1}^{(j)} &= \theta_n^{(j)} + C_n^{\theta w} (C_n^{ww} + \Upsilon_n \Gamma)^{-1} (y_{n+1}^{(j)} - \mathcal{G}(u_n^{(j)})), \end{aligned}$$

where Υ_n is chosen as $\Upsilon_n^{i+1} = 2^i \Upsilon_n^0$,

where Υ_n^0 is an initial guess. We then define $\Upsilon_n \equiv \Upsilon_n^N$ where N is the first integer such that

$$\rho \|\Gamma^{-1/2}(y^{(j)} - \bar{w}_n)\|_{\mathcal{Y}} \leq \Upsilon_n^N \|\Gamma^{1/2}(C_n^{ww} + \Upsilon_n^N \Gamma)^{-1}(y^{(j)} - \bar{w}_n)\|_{\mathcal{Y}}.$$

then we see that

$$u_{n+1}^{(j)} = u_n^{(j)} - \frac{1}{J-1} \sum_{m=1}^J d_n^{(j,m)} u_n^{(m)}, \quad (21)$$

and it is apparent that the algorithm will preserve the linear span of the initial ensemble $\{u_0^{(j)}\}_{j=1}^J$. We describe the details of how Υ_n is chosen in the next subsection where we display the algorithm in full for a generalization of the setting of (1) to the hierarchical setting.

To write down the continuous-time limit of the EKI we consider the setting in which $\Upsilon_n^{-1} \equiv (J-1)h$ and view $u_n^{(j)}$ as approximating a function $u^{(j)}(t)$ at time $t = nh$. If we define

$$d^{(j,m)} = \langle \Gamma^{-1}(\mathcal{G}(u^{(j)}) - y), \mathcal{G}(u^{(m)}) - \bar{\mathcal{G}} \rangle,$$

with obvious definition of $\bar{\mathcal{G}}$, then we obtain the continuous-time limit

$$\dot{u}^{(j)} = - \sum_{m=1}^J d^{(j,m)} u^{(m)}, \quad (22)$$

where $\dot{u}^{(j)}$ denotes the standard time derivative of $u^{(j)}$ viewed as solving an ordinary differential equation; because the algorithm preserves the linear span of the initial ensemble [48] the dynamics take place in a finite dimensional space, even if \mathcal{X} is infinite dimensional. Note that $d^{(j,m)}$ depends on $\{u^{(k)}\}_{k=1}^J$ and that the dynamical system couples the ensemble members.

3.2. Generalization for Centered Hierarchical Inversion

Algorithm 1 shows the generalization of (17) to the setting of centered hierarchical inversion. The ensemble is now over both u and θ and cross covariances from the observational space to both the u and θ spaces are required. Algorithm 1 also spells out in detail how the parameter Υ_n is chosen. We now define

$$d_n^{(j,m)} = \langle (C_n^{ww} + \Upsilon_n \Gamma)^{-1}(\mathcal{G}(u_n^{(j)}) - y), \mathcal{G}(u_n^{(m)}) - \bar{\mathcal{G}}_n \rangle,$$

and we see that

$$u_{n+1}^{(j)} = u_n^{(j)} - \frac{1}{J-1} \sum_{m=1}^J d_n^{(j,m)} u_n^{(m)}, \quad (23)$$

and

$$\theta_{n+1}^{(j)} = \theta_n^{(j)} - \frac{1}{J-1} \sum_{m=1}^J d_n^{(j,m)} \theta_n^{(m)}. \quad (24)$$

It is apparent that, once again, the algorithm will preserve the linear span of the initial ensemble $\{u_0^{(j)}, \theta_0^{(j)}\}_{j=1}^J$. Furthermore we note that for the centered hierarchical method, since \mathcal{G} does not depend on θ , the algorithm projected onto the u coordinate is identical to that in the preceding subsection, with the only difference being that the initial span of $\{u_0^{(j)}\}_{j=1}^J$ is constructed over a diverse set of θ , reflecting the dependency structure

in $\mathbb{P}(u, \theta)$; for hierarchical priors as in subsections 2.2.3 and 2.2.4, the dependency structure is typically of the form $\mathbb{P}(u, \theta) = \mathbb{P}(u|\theta)\mathbb{P}(\theta)$. ¶

If we again define

$$d^{(j,m)} = \langle \Gamma^{-1}(\mathcal{G}(u^{(j)}) - y), \mathcal{G}(u^{(m)}) - \bar{\mathcal{G}} \rangle,$$

with obvious definition of $\bar{\mathcal{G}}$, then we obtain the continuous-time limit

$$\dot{u}^{(j)} = - \sum_{m=1}^J d^{(j,m)} u^{(m)}, \quad (25)$$

$$\dot{\theta}^{(j)} = - \sum_{m=1}^J d^{(j,m)} \theta^{(m)}. \quad (26)$$

Since $d^{(j,m)}$ depends only on $\{u^{(k)}\}_{k=1}^J$ and not on $\{\theta^{(k)}\}_{k=1}^J$ for the centered hierarchical method the continuous time limit for u is identical to that in the preceding subsection, with the only difference being the creation of the initial ensemble using variable θ . This severely limits the capability of the hierarchical method in the centered case, and is motivation for the non-centered approach that we now describe.

3.3. Generalization for Non-Centered Hierarchical Parameterization

One of the reasons for using hierarchical parameterizations is that good choices of parameters such as length scale and regularity of u are not known a priori; this suggests that they might be learnt from the data. In this context the preservation of the linear span of the initial ensemble is problematic if the algorithm is formulated in terms of (u, θ) . This is because, even though the length-scale (for example) may update as the algorithm progresses, the output for u remains in the linear span of the initial set of u , which likely does not contain a good estimate of the true length scale. Instead one can work with the variables (ξ, θ) , where ξ is the forcing function in a stochastic PDE, as explained in subsections 2.2.3 and 2.2.4. Working with (u, θ) and with (ξ, θ) are referred to as the *centered* parameterization and the *non-centered* parameterization, respectively. The pros and cons of each method is discussed in the context of Bayesian inversion in [41, 42], where the terminology is also introduced. The provenance of the terminology has no direct relevance in our context, but we retain it to make the link with the existing literature.

The algorithm for updating (ξ, θ) is identical to that shown in subsection 3.1 with the identifications $u \mapsto (\xi, \theta)$ and $\mathcal{G} \mapsto \mathcal{G} \circ T$. Note that even though, in the centered case, \mathcal{G} depends only on u , the mapping $\mathcal{G} \circ T$ will depend on both ξ and θ . Hence these variables are coupled through the iteration. Indeed if we now define

$$d_n^{(j,m)} = \langle (C_n^{ww} + \Upsilon_n \Gamma)^{-1}(\mathcal{G} \circ T(\xi_n^{(j)}, \theta_n^{(j)}) - y), \mathcal{G} \circ T(\xi_n^{(m)}, \theta_n^{(m)}) - \overline{\mathcal{G} \circ T}_n \rangle$$

¶ The centered hierarchical method, where \mathcal{G} does not depend on θ , is presented for simplicity. The details of this algorithm are readily transferred to include θ dependence in the forward mapping \mathcal{G} as required by the version of the Bayesian level set method advocated in [16].

then we see that

$$\xi_{n+1}^{(j)} = \xi_n^{(j)} - \frac{1}{J-1} \sum_{m=1}^J d_n^{(j,m)} \xi_n^{(m)}, \quad (27)$$

and

$$\theta_{n+1}^{(j)} = \theta_n^{(j)} - \frac{1}{J-1} \sum_{m=1}^J d_n^{(j,m)} \theta_n^{(m)}. \quad (28)$$

Although the algorithm will preserve the linear span of the initial ensemble $\{\xi_0^{(j)}, \theta_0^{(j)}\}_{j=1}^J$ the variable of interest $u_n^{(j)} = T(\xi_n^{(j)}, \theta_n^{(j)})$ is not in the linear span of $u_0^{(j)}$, in general. This confers a significant advantage on the non-centered parameterization in comparison with the centered approach.

As in the previous subsections we describe a continuous-time limit, now for the non-centered hierarchical approach. We define

$$d^{(j,m)} = \langle (\Gamma^{-1}(\mathcal{G} \circ T(\xi^{(j)}, \theta^{(j)}) - y), \mathcal{G} \circ T(\xi^{(m)}, \theta^{(m)}) - \overline{\mathcal{G} \circ T}) \rangle,$$

again with the obvious definition of $\overline{\mathcal{G} \circ T}$. In the same setting adopted in the previous two subsections we obtain the limiting equations

$$\begin{aligned} \dot{\xi}^{(j)} &= - \sum_{m=1}^J d^{(j,m)} \xi^{(m)}, \\ \dot{\theta}^{(j)} &= - \sum_{m=1}^J d^{(j,m)} \theta^{(m)}. \end{aligned}$$

Now $d^{(j,m)}$ depends on both $\{\xi^{(k)}\}_{k=1}^J$ and on $\{\theta^{(k)}\}_{k=1}^J$ so that the dynamical system not only couples the ensemble members but in general can couple the dynamics for ξ and for θ . This is another way to understand the significant advantage of the non-centered parameterization in comparison with the centered approach.

4. Model Problems

In order to demonstrate the benefits of the parameterizations that we introduced here we employ a number of models on which we will base our numerical experiments. This section will be dedicated to describing the various PDEs that will be used. We will describe the forward problem, together with a basic version of the inverse problem, for each model, relevant in the non-hierarchical case. We note that the ideas such as the level set method, and hierarchical formulations from subsections 2.1.2 and 2.1.3, can be used to reformulate the inverse problems, and we will use these reformulations in section 5.

4.1. Model Problem 1

Our first test model is from electrical impedance tomography (EIT). This imaging method is used to learn about interior properties of a medium by injecting current, and

measuring voltages, on the boundary [8, 15]. We will use the complete electrode model (CEM) introduced in [49]. The forward model is as follows: given a domain $D = \mathcal{B}(0, 1)^2$ and a set of electrodes $\{e_l\}_{l=1}^{m_e}$ on the boundary ∂D with contact impedance $\{z_l\}_{l=1}^{m_e}$, and interior conductivity κ , the CEM aims to solve for the potential ν inside the domain D and the voltages $\{V_l\}_{l=1}^{m_e}$ on the boundary. The governing equations are

$$\nabla \cdot (\kappa \nabla \nu) = 0, \quad \in D \quad (29a)$$

$$\nu + z_l \kappa \nabla \nu \cdot n = V_l, \quad \in e_l, \quad l = 1, \dots, m_e \quad (29b)$$

$$\nabla \nu \cdot n = 0, \quad \in \partial D \setminus \cup_{l=1}^{m_e} e_l \quad (29c)$$

$$\int \kappa \nabla \nu \cdot n \, ds = I_l, \quad \in e_l, \quad l = 1, \dots, m_e, \quad (29d)$$

with n denoting the outward normal vector on the boundary. The linearity of the problem implies that the relationship between injected current and measured voltages can be described through an Ohm's Law of the form

$$V = R(\kappa) \times I. \quad (30)$$

In our experiments D will be a two-dimensional disc of radius 1. The inverse problem may now be stated. We write the unknown conductivity as $\kappa = \exp(u)$ and try to infer u from a set of J noisy measurements of voltage/current pairs (V_j, I_j) . If we define $\mathcal{G}_j(u) = R(\kappa) \times I_j$ then the inverse problem is to find u from y given an equation of the form (1).

4.2. Model Problem 2

Our second model problem arises in hydrology: the single-phase Darcy flow equations. The concrete instance of the forward problem is as follows: given the domain $D = [0, 6]^2$ and real-valued permeability function κ defined on D , the forward model is to determine real-valued pressure (or hydraulic head) function p on D from

$$-\nabla \cdot (\kappa \nabla p) = f, \quad x \in D, \quad (31)$$

with mixed boundary conditions

$$p(x_1, 0) = 100, \quad \frac{\partial p}{\partial x_1}(6, x_2) = 0, \quad -\kappa \frac{\partial p}{\partial x_1}(0, x_2) = 500, \quad \frac{\partial p}{\partial x_2}(x_1, 6) = 0, \quad (32)$$

and the source term f defined as

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

The inverse problem concerned with (31) is as follows: write $\kappa = \exp(u)$ and determine u from J linear functionals of the pressure $\mathcal{G}_j(u) = l_j(p)$. This may thus be cast in

the form (1). We take the linear functionals as mollified pointwise observations on a regular grid. This specific set-up of the PDE model is that tested by Hanke [22] in his consideration of the regularized Levenberg-Marquardt algorithm. More information on this setting can be found by Carera et al. [11].

4.3. Model Problem 3

Our final model is a simple linear inverse problem which we can describe directly. The aim is to reconstruct a function u from noisy observation of J linear functionals $\mathcal{G}_j(u) = l_j(p)$, $j = 1, \dots, J$, where p solves the equation

$$\frac{d^2 p}{dx^2} + p = u, \in D, \quad (33a)$$

$$p = 0, \in \partial D. \quad (33b)$$

This may also be cast in the form (1). We use equally spaced pointwise evaluations as our linear functionals. We will assume our domain is chosen such that $D = [0, 10]$.

5. Numerical Examples

To assess the performance of each parameterization we present a range of numerical experiments on each of the three model problems described in the previous section. Our experiments will be presented in a consistent fashion, between the different models and the different algorithms. Each model problem will be tested using each of the non-hierarchical and hierarchical approaches, although we will not use the centered approach for Model Problem 3. Within each of these approaches we will show the progression of the inverse solver from the first to the last iteration. This will include five images ordered by iteration number, with the first figure displaying the first iteration and the last displaying the final iteration. These figures will be accompanied with figures demonstrating the learning of the hyperparameters, as the iteration progresses, for Model Problems 1 and 2, but not for Model Problem 3 (where the hyperparameter is a field).

In order to illustrate the effect of the initial ensemble we will show output of the non-centered approach for ten different initializations, for each model problem. We will plot the final iteration reconstruction arising from four of those initializations. We observe the variation across the initializations through the relative errors in the unknown field u_{EKI} , with respect to the truth u^\dagger , and in the data misfit, as the iteration progresses:

$$\frac{\|u_{\text{EKI}} - u^\dagger\|_{L^2(D)}}{\|u^\dagger\|_{L^2(D)}},$$

$$\|y - \bar{\mathcal{G}}(u_{\text{EKI}})\|_{\Gamma}.$$

5.1. Level Set Parameterization

Level set methods are a computationally effective way to represent piecewise constant functions, and there has been considerable development and application to inverse problems [9, 10, 24], starting from the paper [47], in which interfaces are part of the unknown. We apply level set techniques, combined with hierarchical parameter estimation, in the context of ensemble inversion; we are motivated by the recent Bayesian level set method developed by Lu et al. in [28], and its hierarchical extensions introduced in [14, 16].

When applying the level set technique to inverse problems of the form (1), we modify our forward operator to

$$\mathcal{G} = \mathcal{O} \circ G \circ F, \quad (34)$$

where $G : X \mapsto Y$ maps the coefficient of the PDE to its solution, $\mathcal{O} : Y \rightarrow \mathcal{Y}$ is our observational operator, and $F : \mathcal{X} \rightarrow X$ is the level set map described by

$$(Fu)(x) \rightarrow \kappa(x) = \sum_{i=1}^n \kappa_i \mathbb{1}_{D_i}(x). \quad (35)$$

The sets $\{D_i\}_{i=1}^n$ are n disjoint subdomains with union D and whose boundaries define the interfaces. The boundaries are assumed to be defined through a continuous real-valued function u on D via its level sets. In order to model the level set method hierarchically we will base our reconstructions on the approaches taken in subsection 2.2.4. In general it can be helpful to re-scale the level values as the hierarchical parameter τ is learned [16]; however if the unknown is binary and the level set taken as zero, as used in our numerical experiments here, then this is not a consideration.

We apply level set inversion to Model Problem 1 (EIT) from subsection 4.1. We reconstruct a binary field and the variable u is, rather than the logarithm of the conductivity κ , the level set function defining (35): specifically the level set formulation is achieved through representing the conductivity as

$$\kappa(x) = (Fu)(x) = \kappa_- \chi_{u \leq 0} + \kappa_+ \chi_{u > 0}, \quad (36)$$

where χ_A denotes a characteristic function of A with κ_- and κ_+ being known positive constants that help define low and high levels of our diffusion coefficient.

We place 16 equidistant electrodes on the boundary of the unit disc D in order to define our observations. All experiments are conducted using the MATLAB package EIDORS [3]. The contact impedances $\{z_l\}_{l=1}^{m_e}$ are chosen with value 0.05 and all electrodes chosen subjected to an input current of 0.1. This provides a matrix of stimulation patterns $I = \{I^{(j)}\}_{j=1}^{15} \in \mathbb{R}^{16 \times 15}$ given as

$$I = 0.1 \times \begin{pmatrix} +1 & 0 & \dots & 0 \\ -1 & +1 & \dots & 0 \\ 0 & -1 & \ddots & 0 \\ \vdots & \vdots & \ddots & +1 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

For our iterative method we choose $J = 200$ ensemble members with regularization parameter $\rho = 0.8$. The covariance of our noise η is chosen such that $\Gamma = 10^{-4} \times I$. Our truth for the EIT problem will take the form given in Figure 5 where we have high levels of conductivity within the two inclusions. This is constructed by thresholding a Whittle-Matérn Gaussian random field defined by (12), (13); true values for the hierarchical parameters used are shown in Table 1.

Remark 5.1 *We do not display the underlying Gaussian random field u which is thresholded to obtain the true conductivity in Figure 5 as this Gaussian random field cannot be expected to be reconstructed accurately, in general. Furthermore it is important to appreciate that in general a true conductivity will not be constructed by such thresholding; the field u is simply an algorithmic construct. We do however show u , and its evolution, in the algorithm, because this information highlights the roles of the length scale and regularity parameters.*

When performing inversion we sample initial ensembles using the prior distributions shown in Table 2. We have set our prior distributions in such a way that the true value for each hyperparameter lies within the range specified.

Hyperparameter	Value
α^\dagger	3
τ^\dagger	10

Table 1. Model problem 1. True values for each hyperparameter.

Hyperparameter	Prior
α	$\mathcal{U}[1.3, 4]$
τ	$\mathcal{U}[5, 30]$

Table 2. Model problem 1. Prior distribution for each hyperparameter.

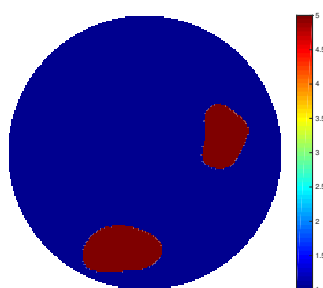


Figure 5. Model problem 1: true log-conductivity.

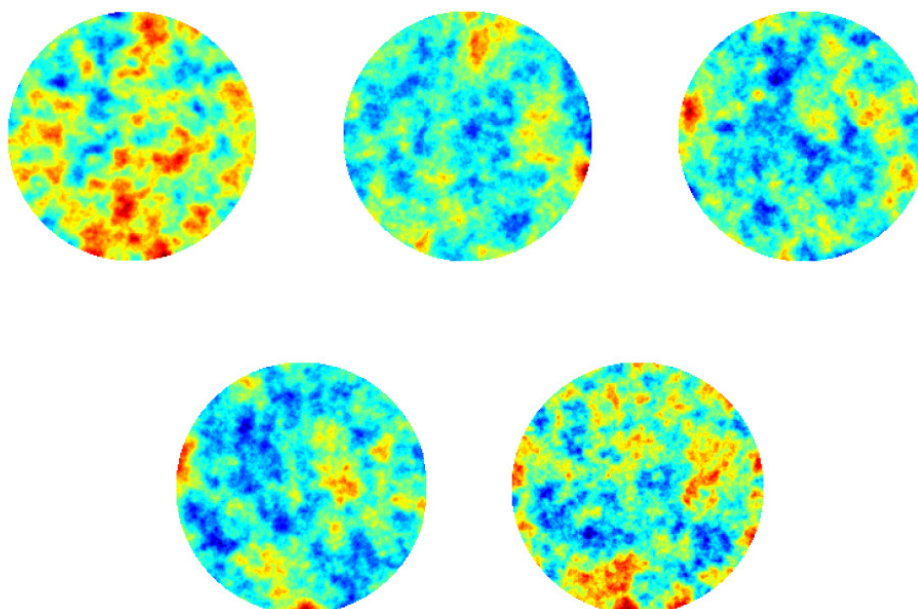


Figure 6. Model problem 1. Progression through iterations of non-hierarchical method.

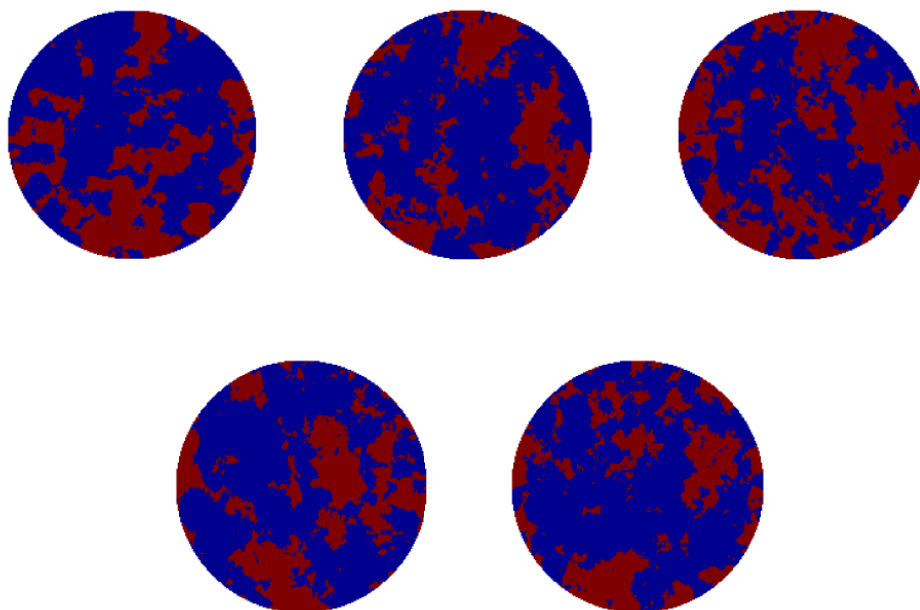


Figure 7. Model problem 1. Progression through iterations of non-hierarchical method with level set.

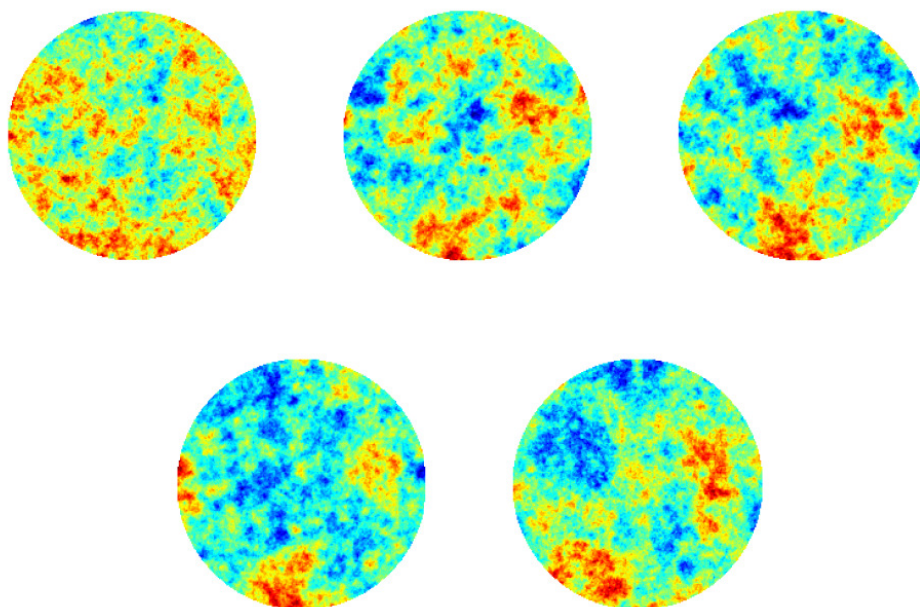


Figure 8. Model problem 1. Progression through iterations of centered hierarchical method.

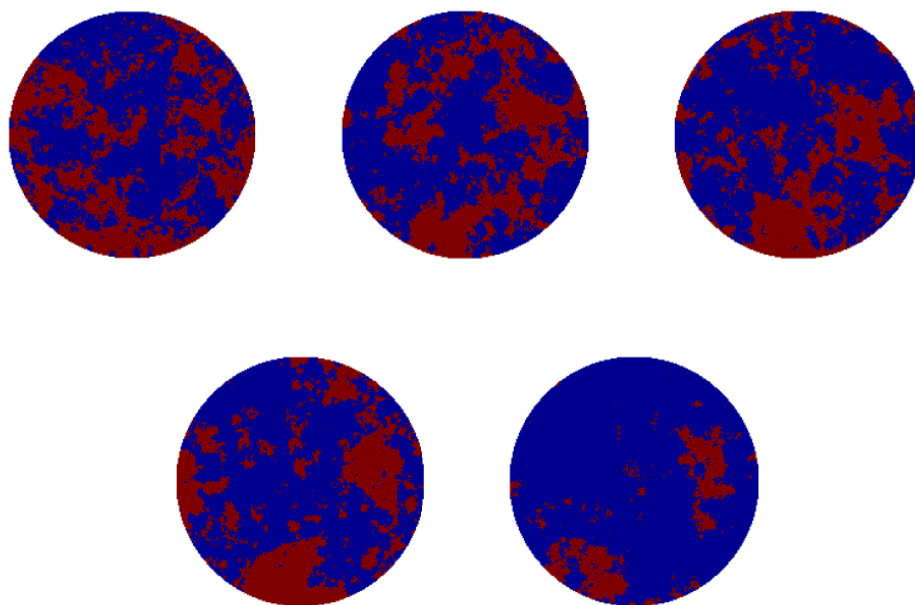


Figure 9. Model problem 1. Progression through iterations of centered hierarchical method with level set.

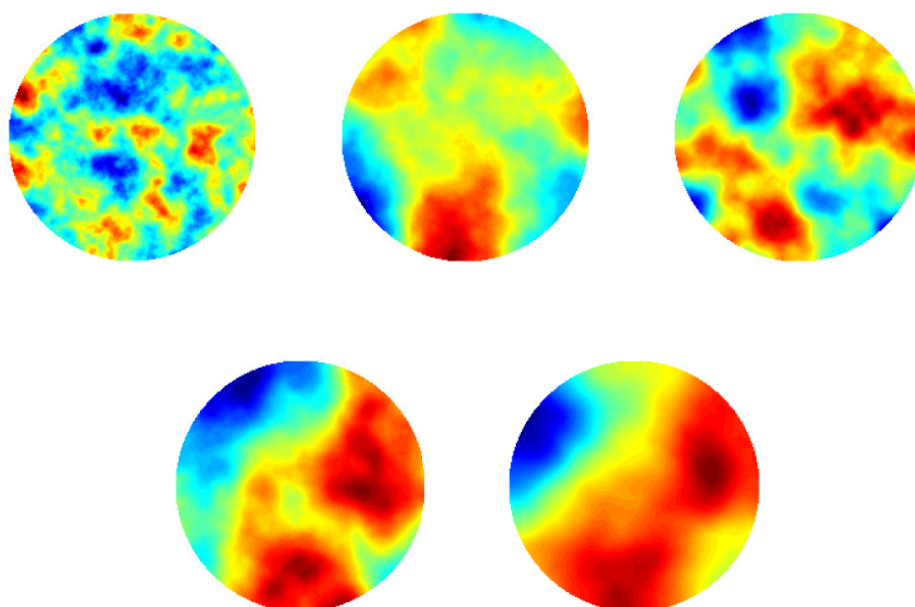


Figure 10. Model problem 1. Progression through iterations of non-centered hierarchical method.

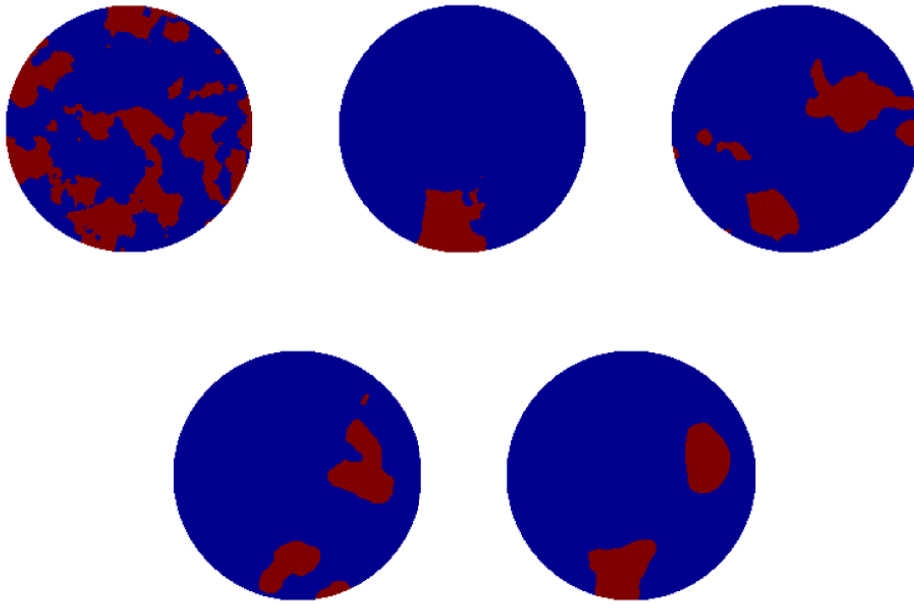


Figure 11. Model problem 1. Progression through iterations of non-centered method with level set.

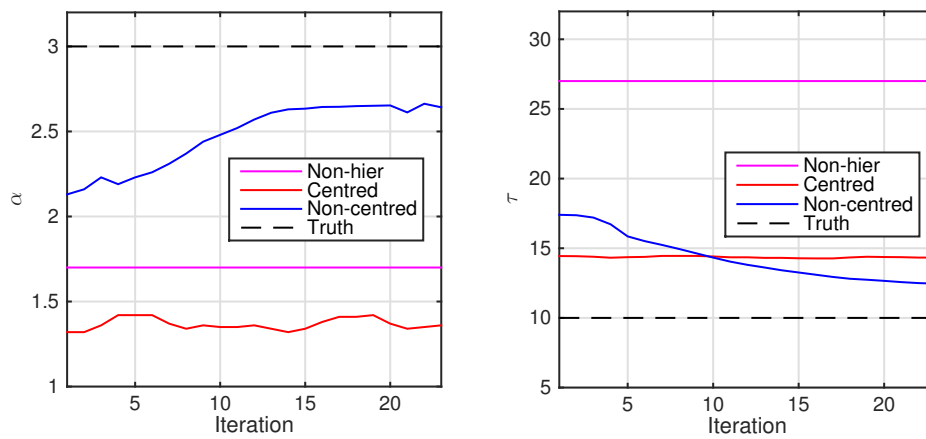


Figure 12. Model problem 1. Progression of average value for α and τ .

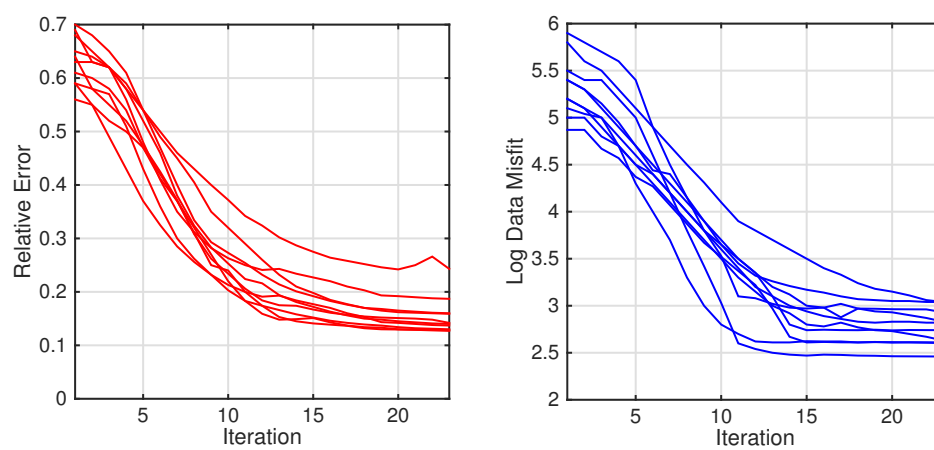


Figure 13. Model problem 1. Left: relative error. Right: log-data misfit.

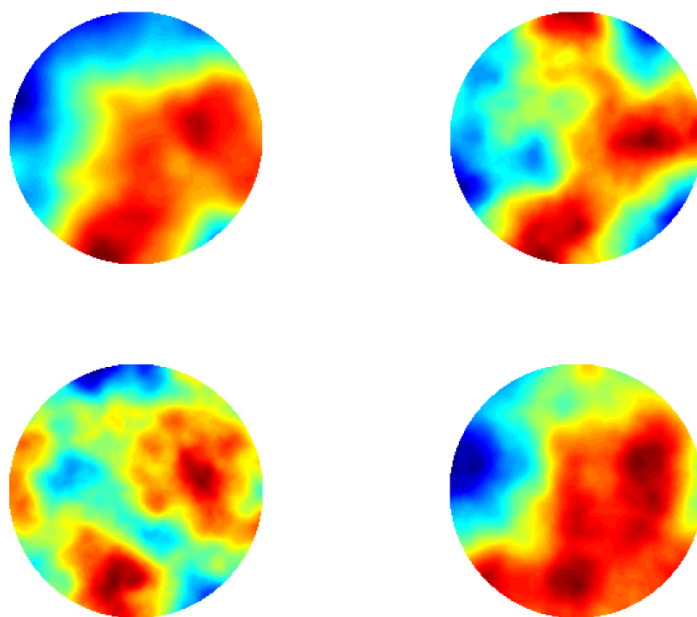


Figure 14. Model problem 1. EKI for the final iteration for the non-centered approach with Whittle-Matérn from four different initializations.

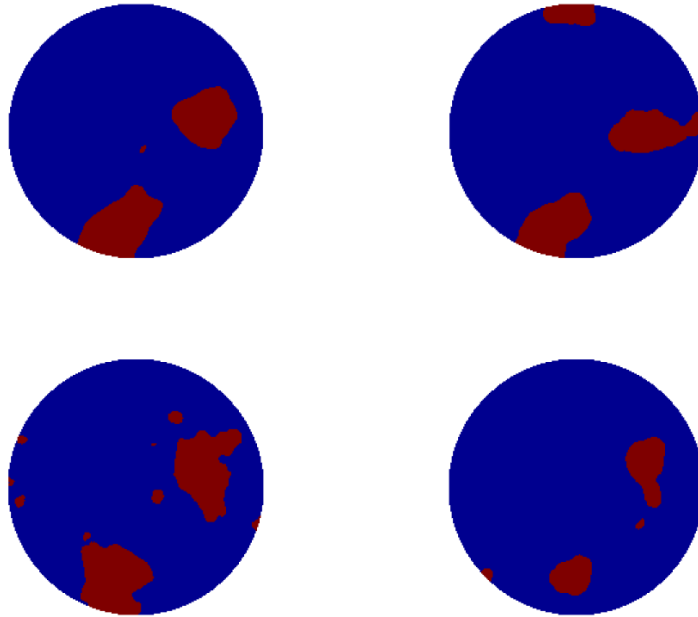


Figure 15. Model problem 1. EKI for the final iteration for the non-centered method with level set from four different initializations.

Figures 6 - 11 show the progression of both the level-set function u , and the permeability κ , through five iterations of the method. Non-hierarchical, centered and non-centered methods are considered in turn. By comparing the reconstructions with the true conductivity, these figures clearly demonstrate two facts: (a) that being hierarchical is necessary to obtain a good reconstruction; (b) that implementing the hierarchical method using a non-centered parameterization has significant benefits when compared to the centered method. These points are further demonstrated in Figures 12 which shows the learning of the hyperparameters, in comparison with the truth, for the three methods.

Concentrating solely on the non-centered approach, we run ten different initializations of the EKI. We display the resulting data-misfit and error, as a function of iteration, for all ten in Figure 13. We display the last iteration of four of these ten in Figures 14 - 15.

In summary, Figures 10 - 12 clearly show the superiority of the non-centered hierarchical method. For all the initializations shown, the EKI produces conductivities which concentrate near to the true conductivity and have length scale similar to those appearing in the truth; see Figures 13 - 15. The centered hierarchical and non-hierarchical methods fail to do this; see Figures 6 - 9. However it is important to note the non-centered method does produce substantial variation in the predicted solution, depending on which initialization is used as shown in Figures 13 - 15.

5.2. Geometric Parameterization

In this subsection we employ Model Problem 2 from subsection 4.2. We consider reconstruction of a piecewise continuous channel which is defined through two heterogeneous Gaussian random fields, scalar geometric parameters specifying the geometry and scalar hierarchical parameters characterizing the length-scale and regularity of the two fields.

The truth u^\dagger is shown in Figure 16. It is drawn from a prior distribution which we now describe; details may be found in [27]. The channel is described by five parameters: d_1 – amplitude; d_2 – frequency; d_3 – angle; d_4 – initial point; and d_5 – width. We generate two Gaussian random fields $\{\kappa_i\}_{i=1}^2$, both defined on the whole of the domain D but entering the permeability κ only inside and outside (respectively) the channel. The unknown u thus comprises the five scalars $\{d_i\}_{i=1}^5$ and the two fields $\{\kappa_i\}_{i=1}^2$. We do not explicitly spell out the mapping from u to the coefficient κ appearing in the Darcy flow, but leave this to the reader. The $\{\kappa_i\}_{i=1}^2$ are specified as log-normal random fields and the underlying Gaussians are of Whittle-Matérn type, defined by (12), (13); different uniform distributions on α and τ are used for the two fields $\{\kappa_i\}_{i=1}^2$. The parameters $\{d_i\}_{i=1}^5$ are also given uniform distributions. The entire specification of the prior is given in Table 3. For the truth the true hierarchical parameters are provided in Table 4.

In our inversion we employ 64 mollified pointwise observations $\{l_i(p)\}_{i=1}^{64}$ given by, for some $\sigma > 0$,

$$l_t(p) = \int_D \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2\sigma^2}(x-x_i)^2} p(x) dx, \quad (37)$$

where the x_i are uniformly distributed points on D . We discretize the forward model using a second order centered finite difference method with mesh spacing 10^{-2} . For our EKI method we use the same values for our parameters as in subsection 5.1.

Parameter	Prior
d_1	$\mathcal{U}[0, 1]$
d_2	$\mathcal{U}[2, 13]$
d_3	$\mathcal{U}[0.4, 1]$
d_4	$\mathcal{U}[0, 1]$
d_5	$\mathcal{U}[0.1, 0.3]$
κ_1	$N(1, (I - \tau_1^2 \Delta)^{-\alpha_1})$
κ_2	$N(4, (I - \tau_2^2 \Delta)^{-\alpha_2})$
α_1	$\mathcal{U}[1.3, 3]$
τ_1	$\mathcal{U}[8, 30]$
α_2	$\mathcal{U}[1.3, 3]$
τ_2	$\mathcal{U}[8, 30]$

Table 3. Model problem 2. Prior associated with channelised flow.

Parameter	Value
α_1^\dagger	2
α_2^\dagger	2.8
τ_1^\dagger	30
τ_2^\dagger	10

Table 4. Model problem 2. Parameter selection of the truth.

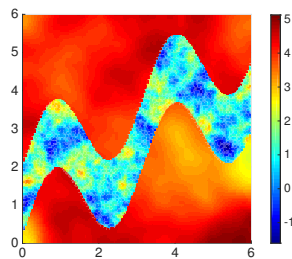


Figure 16. Model problem 2. True log-permeability.

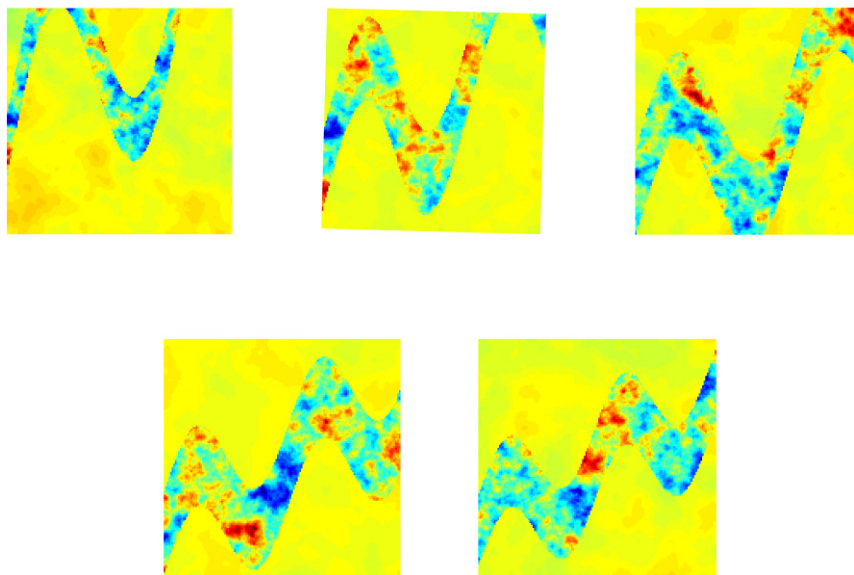


Figure 17. Model problem 2. Progression through iterations of non-hierarchical method.

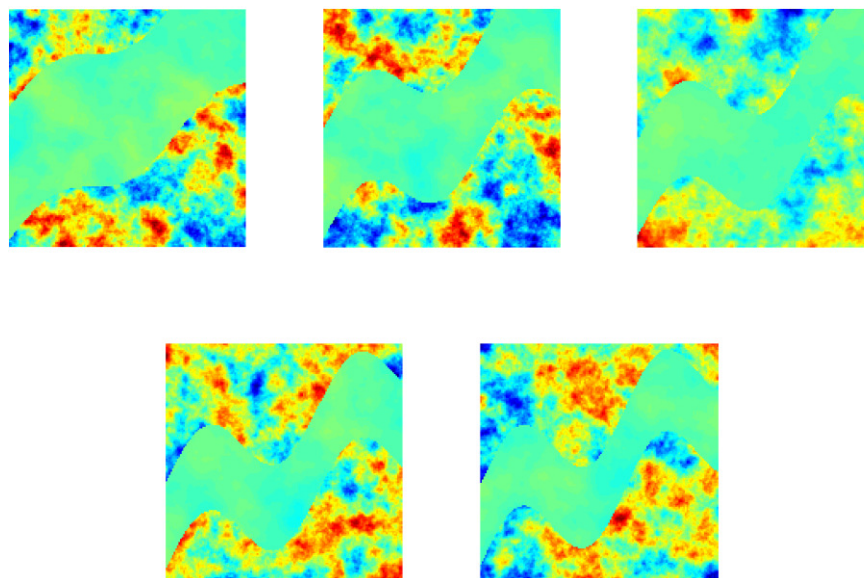


Figure 18. Model problem 2. Progression through iterations of centered hierarchical method.

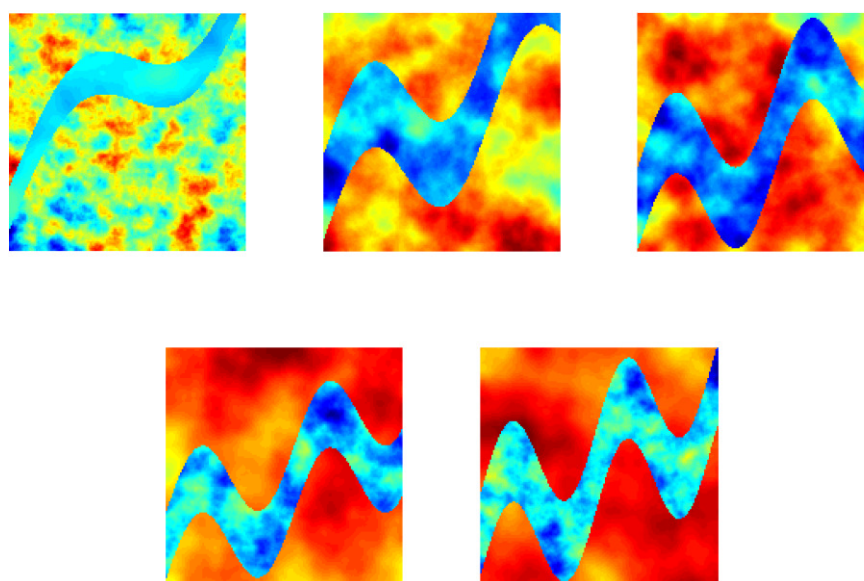


Figure 19. Model problem 2. Progression through iterations of non-centered hierarchical method.

Figures 17 - 19 are consistent with the previous subsection in that we notice that hierarchical methods are needed and that non-centring is necessary to make hierarchical

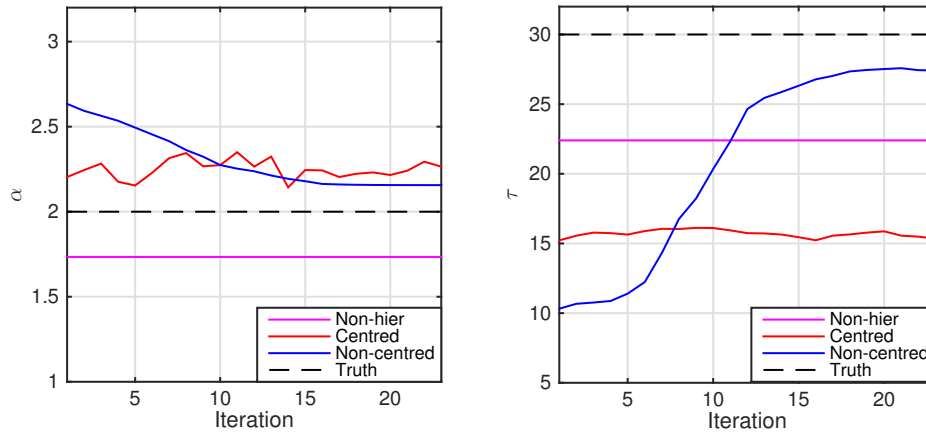


Figure 20. Model problem 2. Progression of average value for α_1 and τ_1 .

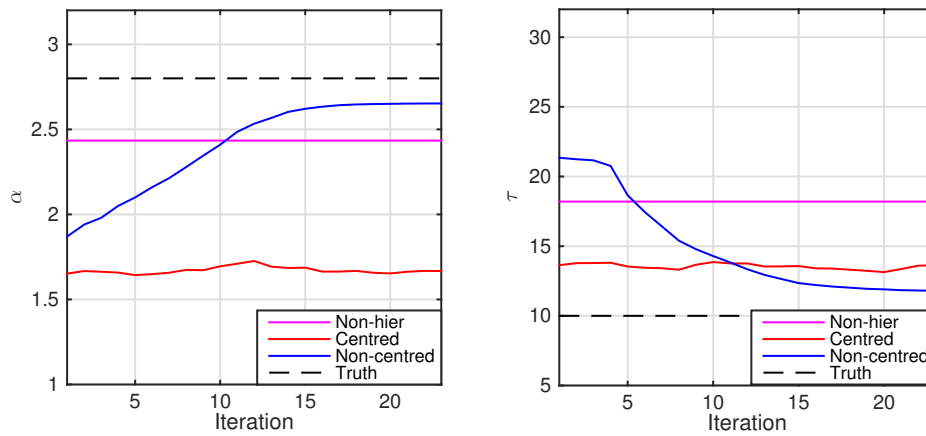


Figure 21. Model problem 2. Progression of average value for α_2 and τ_2 .

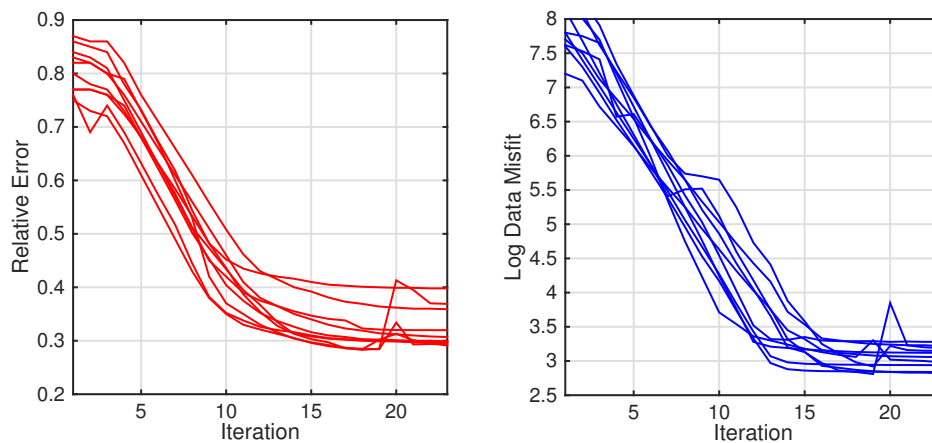


Figure 22. Model problem 2. Left: relative error. Right: log-data misfit.

methods perform well. We qualify this by noting that the geometry is well-learned in all cases, but that the reconstructions of the random fields u_1 and u_2 inside and outside the geometry are sensitive to needing non-centered hierarchical representation. Even then the reconstruction is only accurate in terms of amplitude and length scales and not pointwise.

Figures 20 and 21 give further insight into this, showing how the smoothness and length scale parameters are learned differently in the hierarchical, centered and non-centered methods. Again the conclusions are consistent with the previous subsection. Figures 22 and 23 concentrate on the application of the non-centered approach, using ten different initializations. The data misfit and relative error in the field are shown for all ten cases in Figure 22; four solution estimates are displayed in Figure 23. The results are similar to those in the previous subsection. However there is more variability across initializations. This might be ameliorated by use of a larger ensemble size.

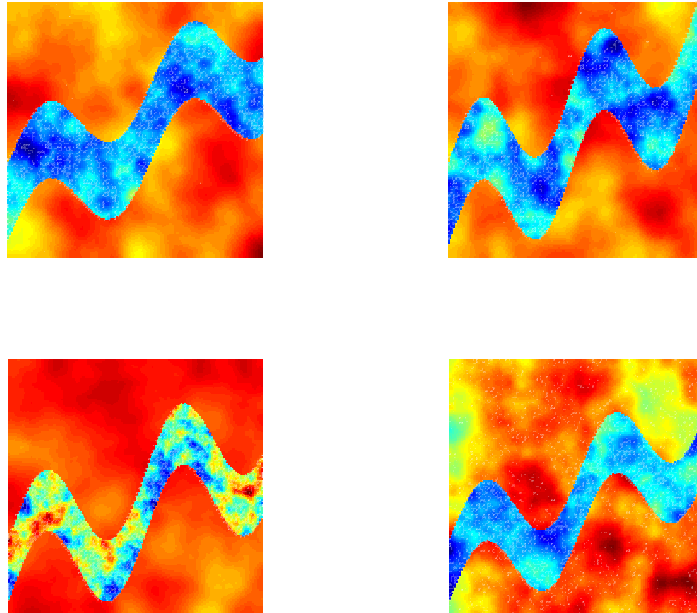


Figure 23. Model problem 2. EKI for the final iteration for the non-centered approach from four different initializations.

5.3. Function-valued Hierarchical Parameterization

Our final set of experiments will be based on hierarchical inversion of non-stationary random fields, using Model Problem 3. We consider reconstruction of truths that are not drawn from the prior; the prior will be a hierarchical Gaussian model with spatially varying inverse length scale as hyperparameter. Examples of such truths are ones which contain both rough and smooth features. We discretize the forward model using a piecewise-linear finite element method (FEM), with a mesh of $h = 1/100$. The truth we aim to recover is given by

$$u^\dagger(x) = \begin{cases} \exp\left(4 - \frac{25}{x(5-x)}\right), & x \in (0, 5) \\ 1, & x \in [7, 8] \\ -1, & x \in (8, 9) \\ 0, & \text{otherwise,} \end{cases} \quad (38)$$

which incorporates both rough and smooth features. Our parameters for the iterative methods are identical to those used in subsection 5.1. Because the results of subsection 5.1 and 5.2 clearly demonstrate the need for non-centring in hierarchical methods, we do not include results for the centered hierarchical approach here; we compare non-hierarchical methods with the use of non-centered hierarchical methods with both Cauchy and Gaussian random fields as priors on the hyperparameter v .

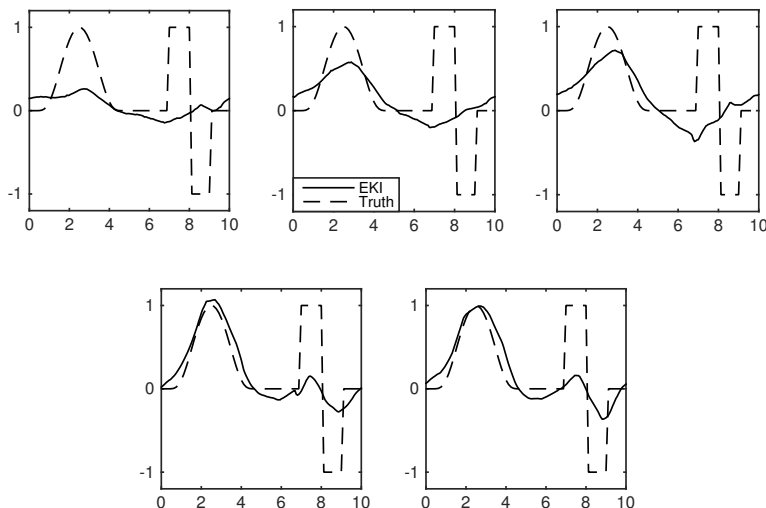


Figure 24. Model problem 3. Progression through iterations of non-hierarchical method.

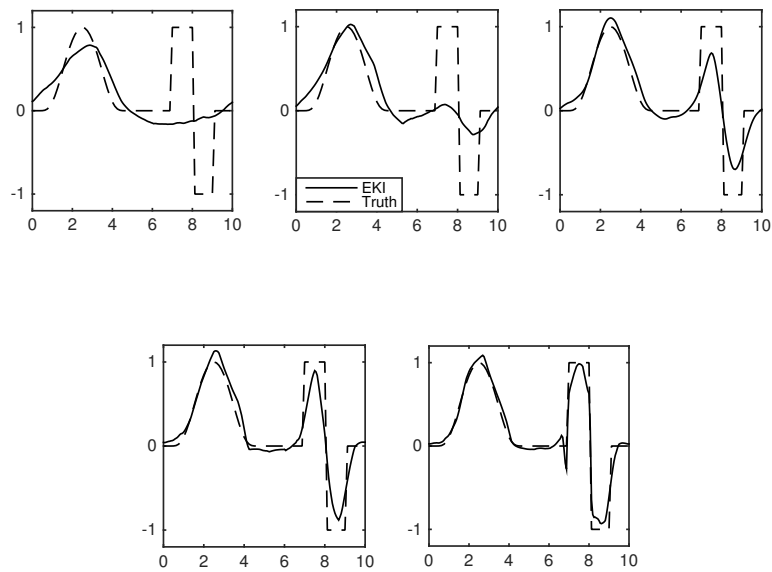


Figure 25. Model problem 3. Progression through iterations with hierarchical Gaussian random field.

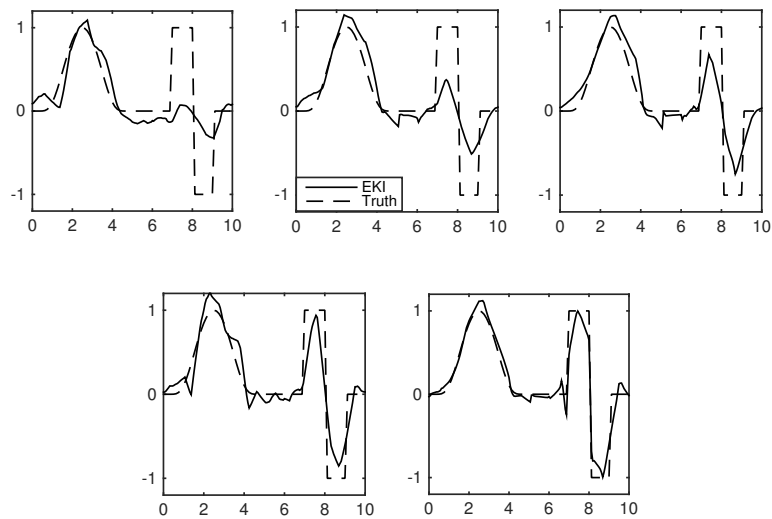


Figure 26. Model problem 3. Progression through iterations with hierarchical Cauchy random field.

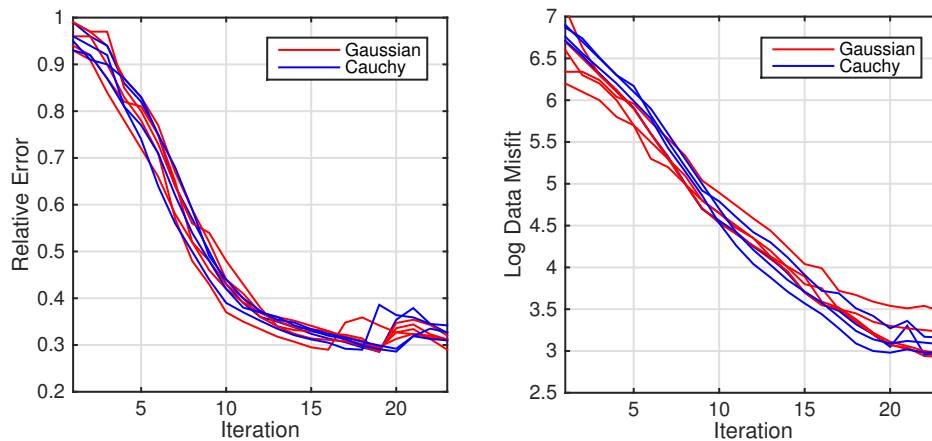


Figure 27. Model problem 3. Left: relative error. Right: log-data misfit.

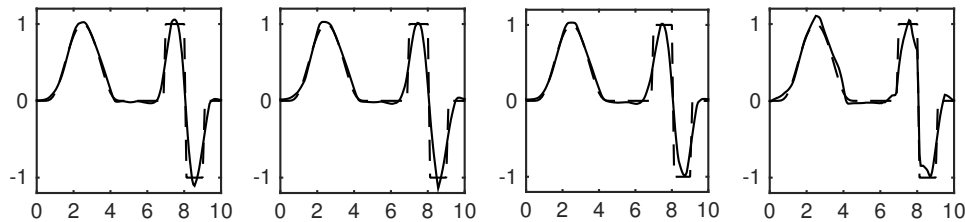


Figure 28. Model problem 3. EKI for the final iteration for the Gaussian hierarchical method from four different initializations.

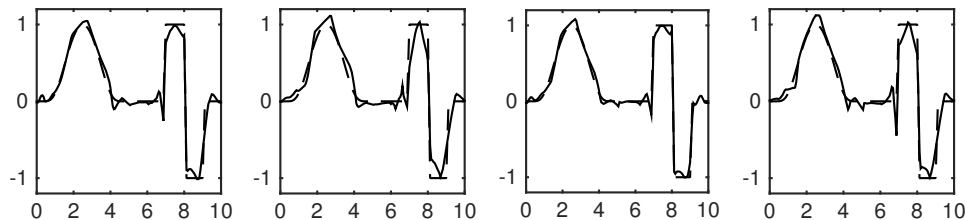


Figure 29. Model problem 3. EKI for the final iteration for the Cauchy hierarchical method from four different initializations.

For our Cauchy density function (16), we set $a = 4$ and $b = d = 0$. Our length scale for the non-hierarchical method will be based on a Gaussian random field (15), similarly with the non-centered Gaussian approach. Our comparison of each approach is provided in Figures 24 - 26. We notice that the non-hierarchical method struggles to reconstruct the truth, and in particular the piecewise constant part of it with discontinuities. In contrast both non-centered approaches perform well. The effectiveness of the non-centered approaches are highlighted in Figures 27 - 29. The first shows the data-misfit and relative error over ten realizations, and the second shows four reconstructions chosen from these ten at random. Substantial robustness to the choice of realization is clear. Note however that for a multi-dimensional problem inversion the variability with respect

to initialization may be more significant, as in the previous two subsections.

6. Conclusion & Discussion

In this paper we have considered several forms of parameterizations for EKI. In particular our main contribution has been to highlight the potential for the use of hierarchical techniques from computational statistics, and the use of geometric parameterizations, such as the level set method. Our perspective on EKI is that it forms a derivative free optimizer and we do not evaluate it from the perspective of uncertainty quantification. However the hierarchical and level set ideas are motivated by Bayesian formulations of inverse problems. We have shown that our parameterizations do indeed lead to better reconstructions of the truth on a variety of model problems including groundwater flow, EIT and source inversion. There is very little analysis of EKI, especially in the fixed, small, ensemble size setting where it is most powerful. Existing work in this direction may be found in [7, 48]; it would be of interest to extend these analyses to the parameterizations introduced here. Furthermore, from a practical perspective, it would be interesting to extend the deployment of the methods introduced here to the study of further applications.

Acknowledgements. The authors are grateful to M. M. Dunlop, O. Papaspiliopoulos and G. O. Roberts for helpful discussions about centering and non-centering parameterizations. The research of AMS was partially supported by the EPSRC programme grant EQUIP and by AFOSR Grant FA9550-17-1-0185. LR was supported by the EPSRC programme grant EQUIP. NKC was partially supported by the EPSRC MASDOC Graduate Training Program and by Premier Oil.

References

- [1] Aanonsen S I, Naevdal G, Oliver D S, Reynolds A C and Valles B 2009 The Ensemble Kalman Filter in Reservoir Engineering—a Review *SPE J.* **14** 393-412.
- [2] Anderson J L 2001 An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review* **129** 2884.
- [3] Adler A and Lionheart W R B 2006 Uses and abuses of EIDORS: An extensible software base for EIT *Physiol Meas* **27** 25-42.
- [4] Agapiou S, Bardsley J M, Papaspiliopoulos O and Stuart A M 2014 Analysis of the Gibbs sampler for hierarchical inverse problems. *SIAM Journal on Uncertainty Quantification* **1** 511-544.
- [5] Bakushinskyy A B and Kokurin M Y 2004 Iterative Methods for Approximate Solution of Inverse Problems *Mathematics and its applications*, Springer.
- [6] Bauer F, Hohage T and Munk A 2009 Iteratively Regularized Gauss-Newton Method for Nonlinear Inverse Problems with Random Noise *SIAM J. Numer. Anal* **47** 1827-1846.
- [7] Blömker D, Schillings C and Wacker P 2017 A strongly convergent numerical scheme from EnKF continuum analysis arXiv:1703.06767.
- [8] Borcea L 2002 Electrical Impedance Tomography *Inverse Problems* **18** 99-136.
- [9] Burger M 2002 A framework for the construction of level set methods for shape optimization and reconstruction *Interfaces and Free Boundaries* **5** 301-329.

- [10] Burger M and Osher S 2005 A survey on level set methods for inverse problems and optimal design *Eur. J. Appl. Math* **16** 263-301.
- [11] Carrera J and Neuman S P 1986 Estimation of aquifer parameters under transient and steady state conditions: application to synthetic and field data *Water Resources Research* **22** 228-242.
- [12] Chada N K 2018 Ensemble Based Methods for Geometric Inverse Problems *PhD Dissertation*.
- [13] Dashti M Stuart A M 2017 The Bayesian Approach To Inverse Problems *Handbook of Uncertainty Quantification* 311-428.
- [14] Chen V, Dunlop M M, Papaspiliopoulos O and Stuart A M 2017 Robust MCMC Sampling with Non-Gaussian and Hierarchical Priors in High Dimensions, *In preparation*.
- [15] Dunlop M M and Stuart A M 2016 The Bayesian formulation of EIT *Inverse Problems and Imaging* **10** 1007-1036.
- [16] Dunlop M M, Iglesias M A and Stuart A M 2016 Hierarchical Bayesian Level Set Inversion *Statistics and Computing*.
- [17] Emerick A A 2016 Towards a hierarchical parametrisation to address prior uncertainty in ensemble-based data assimilation *Computational Geosciences* **20** 35-47.
- [18] Ernst, O and Sprungk, B and Starkloff, H. 2015 Analysis of the ensemble and polynomial chaos Kalman filters in Bayesian inverse problems, *SIAM-ASA JUQ* **3** 823-851.
- [19] Evensen 1994 G Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics *J. Geophysical Research-All Series* **99** 10-10.
- [20] Evensen G 2009 *Data Assimilation: The Ensemble Kalman Filter* Springer.
- [21] Evensen G and Van Leeuwen P J 1996 Assimilation of geosat altimeter data for the agulhas current using the ensemble Kalman filter with a quasi-geostrophic model *Monthly Weather Review* **128** 85-96.
- [22] Hanke M 1997 A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems *Inverse Problems* **13** 79-95.
- [23] Houtekamer P L and Mitchell H L 2001 A sequential ensemble Kalman filter for atmospheric data assimilation *Monthly Weather Review* **129** 123-137.
- [24] Iglesias M A 2016 A regularising iterative ensemble Kalman method for PDE-constrained inverse problems *Inverse Problems* **32** .
- [25] Iglesias M A, Law K J H and Stuart A M 2013 Evaluation of Gaussian approximations for data assimilation in reservoir models *Computational Geosciences*, Volume **17** 851-885.
- [26] Iglesias M A, Law K J H and Stuart A M 2014 Ensemble Kalman methods for inverse problems *Inverse Problems* **29**.
- [27] Iglesias M A, Lin K, and Stuart A M 2014 Well-posed Bayesian geometric inverse problems arising in subsurface flow **30**.
- [28] Iglesias M A, Lu Y and Stuart A M 2015 A Bayesian Level Set Method for Geometric Inverse Problems *Interfaces and Free Boundary Problems*.
- [29] Kaipio J and Somersalo E 2004 *Statistical and Computational Inverse Problems* Applied Mathematical Sciences, Springer.
- [30] Kalman R E 1960 A new approach to linear filtering and prediction problems *Trans ASME (J. Basic Engineering)* **82** 35-45.
- [31] Kantas N, Beskos A and Jasra A 2014 Sequential Monte Carlo Methods for High-Dimensional Inverse Problems: A case study for the Navier-Stokes equations *SIAM/ASA J. Uncertain. Quantif* **2** 464-489.
- [32] Li G and Reynolds A C 2009 Iterative ensemble Kalman filters for data assimilation *SPE J* **14** 496-505.
- [33] Majda A and Wang W 2006 *Non-linear Dynamics and Statistical Theories for Basic Geophysical Flows* Cambridge University Press.
- [34] Markkanen M, Roininen L, Huttunen J M J and Lasanen S 2015 Cauchy difference priors for edge-preserving Bayesian inversion with an application to X-ray tomography arXiv:1603.06135.
- [35] Law K J H and Stuart A M 2012 Evaluating Data Assimilation Algorithms *Mon. Weather Rev*

- 140 37-57.
- [36] Liang P, Jordan M I and Klein D 2010 Learning Programs: A Hierarchical Bayesian Approach *Conference: Proceedings of the 27th International Conference on Machine Learning*.
 - [37] Lindgren F, Rue H and Lindström J 2011 An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach **73** 423-498.
 - [38] Liu W, Li J and Marzouk Y M 2017 An approximate empirical Bayesian method for large-scale linear-Gaussian inverse problems arXiv:1705.07646.
 - [39] Myserth I and Omre H 2010 Hierarchical Ensemble Kalman Filter *SPE J* **15**.
 - [40] Oliver D, Reynolds A C and Liu N 2008 *Inverse Theory for Petroleum Reservoir Characterization and History Matching* Cambridge University Press.
 - [41] Papaspiliopoulos O, Roberts G O and Sköld M 2003 Non-centered parameterisations for hierarchical models and data augmentation In *Bayesian Statistics 7: Proceedings of the Seventh Valencia International Meeting*. Oxford University Press.
 - [42] Papaspiliopoulos O, Roberts G O and Sköld M 2007 A general framework for the parametrization of hierarchical models *Statistical Science* 59-73.
 - [43] Pierre Del Moral A J and Doucet A 2006 Sequential Monte Carlo samplers *Journal of the Royal Statistical Society Series B* 411-436.
 - [44] Roininen L, Girolami M, Lasanen S and Markkanen M 2017 Hyperpriors for Matrn fields with applications in Bayesian inversion arXiv:1612.02989.
 - [45] Roininen L, Huttunen J M J and Lasanen S 2014 Whittle-Matérn priors for Bayesian statistical inversion with applications in electrical impedance tomography *Inverse problems and imaging* **8** .561-586.
 - [46] Ruggeri F, Sawlan Z, Scavino M and Tempone R 2016 A Hierarchical Bayesian Setting for an Inverse Problem in Linear Parabolic PDEs with Noisy Boundary Conditions *Bayesian Anal.*
 - [47] Santosa, F 1996 A level-set approach for inverse problems involving obstacles *ESAIM*, 1(January):17-33, 1996.
 - [48] Schillings C and Stuart A M 2017 Analysis of the ensemble Kalman filter for inverse problems. *SIAM J Numerical Analysis* **55**, 1264-1290.
 - [49] Somersalo E, Cheney M and Isaacson D 1992 Existence and Uniqueness for Electrode Models for Electric Current Computed Tomography *SIAM J. Appl. Math* **52** 023-1040
 - [50] Teh Y W, Jordan M I, Beal M J, and Blei D M 2016 Hierarchical Dirichlet processes *Journal of the American Statistical Association* **101** 1566-1581.
 - [51] Tsyrlunikov M and Rakitko A 2016 Hierarchical Bayes Ensemble Kalman Filter *Physica D: Nonlinear Phenomena*.
 - [52] Tarantola A 1987 *Inverse Problem Theory and Methods for Model Parameter Estimation* Elsevier.
 - [53] Tippett M K 2002 Ensemble Square Root Filters *Monthly Weather Review* **131**.