

Hierarchical Bayesian level set inversion

Matthew M. Dunlop¹  · Marco A. Iglesias² · Andrew M. Stuart¹

Received: 13 January 2016 / Accepted: 9 September 2016
© Springer Science+Business Media New York 2016

Abstract The level set approach has proven widely successful in the study of inverse problems for interfaces, since its systematic development in the 1990s. Recently it has been employed in the context of Bayesian inversion, allowing for the quantification of uncertainty within the reconstruction of interfaces. However, the Bayesian approach is very sensitive to the length and amplitude scales in the prior probabilistic model. This paper demonstrates how the scale-sensitivity can be circumvented by means of a hierarchical approach, using a single scalar parameter. Together with careful consideration of the development of algorithms which encode probability measure equivalences as the hierarchical parameter is varied, this leads to well-defined Gibbs-based MCMC methods found by alternating Metropolis–Hastings updates of the level set function and the hierarchical parameter. These methods demonstrably outperform non-hierarchical Bayesian level set methods.

Keywords Inverse problems for interfaces · Level set inversion · Hierarchical Bayesian methods

Electronic supplementary material The online version of this article (doi:[10.1007/s11222-016-9704-8](https://doi.org/10.1007/s11222-016-9704-8)) contains supplementary material, which is available to authorized users.

✉ Matthew M. Dunlop
mdunlop@caltech.edu

¹ Computing & Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125, USA

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

1 Introduction

1.1 Background

The level set method has been pervasive as a tool for the study of interface problems since its introduction in the 1980s Osher and Sethian (1988). In a seminal paper in the 1990s, Santosa demonstrated the power of the approach for the study of inverse problems with unknown interfaces Santosa (1996). The key benefit of adopting the level set parametrization of interfaces is that topological changes are permitted. In particular for inverse problems the number of connected components of the field does not need to be known a priori. The idea is illustrated in Fig. 1. The type of unknown functions that we might wish to reconstruct are piecewise continuous functions, illustrated in the bottom row by piecewise constant ternary functions. However in the inversion, we work with a smooth function, shown in the top row and known as the *level set function*, which is thresholded to create the desired unknown function in the bottom row. This allows the inversion to be performed on smooth functions, and allows for topological changes to be detected during the course of algorithms. After Santosa's paper there were many subsequent papers employing the level set representation for classical inversion, and examples include (Burger 2001; Tai and Chan 2004; Chung et al. 2005; Dorn and Lesselier 2006), and the references therein.

In many inverse problems arising in modern day science and engineering, the data are noisy and prior regularizing information is naturally expressed probabilistically since it contains uncertainties. In this context, Bayesian inversion is a very attractive conceptual approach Jari (2005). Early adoption of the Bayesian approach within level set inversion, especially in the context of history matching for reservoir simulation, includes the papers Xie et al. (2011), Ping

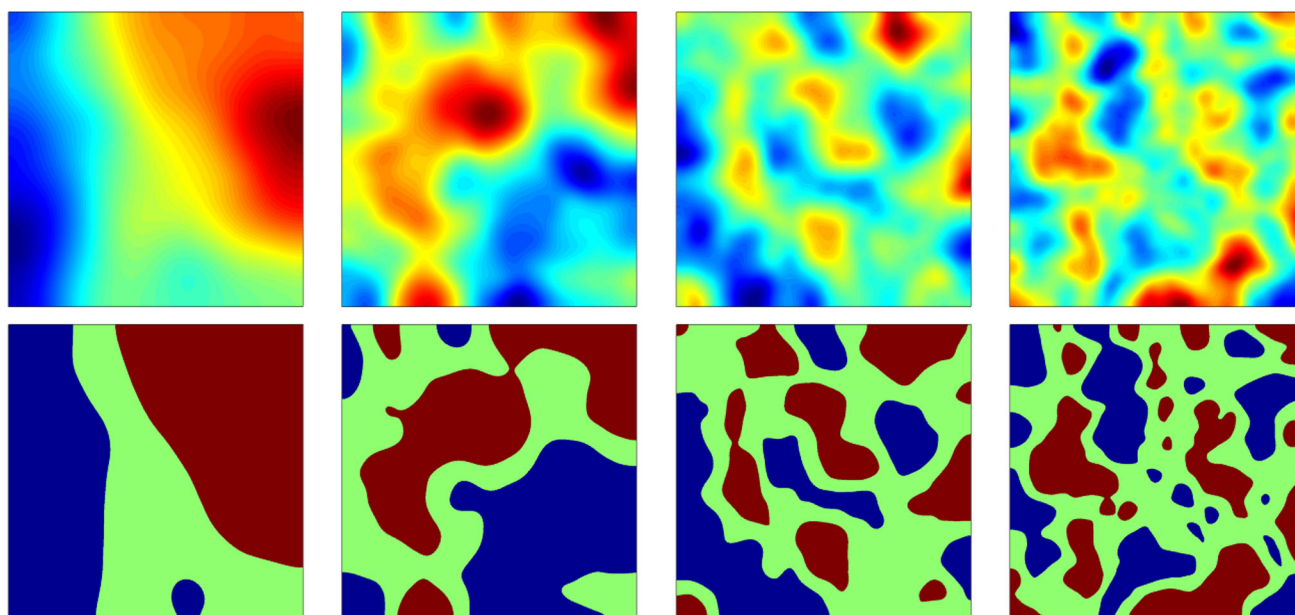


Fig. 1 Four continuous scalar fields (*top*) and the corresponding ternary fields formed by thresholding these fields at two levels (*bottom*). The smooth function in the *top* row is known as the *level set function*

and is used in the inversion procedure. The discontinuous function in the *bottom* row is the physical unknown

and Zhang (2014), Lorentzen et al. (2012), and Lorentzen et al. (2012). In a recent paper Iglesias et al. (2016) the mathematical foundations of Bayesian level set inversion were developed, and a well-posedness theorem established, using the infinite-dimensional Bayesian framework developed in Stuart (2010), Lasanen (2012), Lasanen (2012), and Dashti and Stuart (2016). An ensemble Kalman filter method has also been applied in the Bayesian level set setting Iglesias (2016) to produce estimates of piecewise constant permeabilities/conductivities in groundwater flow/electrical impedance tomography (EIT) models.

For linear Bayesian inverse problems, the adoption of Gaussian priors leads to Gaussian posteriors, formulae for which can be explicitly computed (Franklin 1970; Lehtinen et al. 1999; Mandelbaum 1984). However the *level set map*, which takes the smooth underlying level set function (top row, Fig. 1) into the physical unknown function (bottom row, Fig. 1), is non-linear; indeed it is discontinuous. As a consequence, Bayesian level set inversion, even for inverse problems which are classically speaking ‘linear,’ does not typically admit closed form solutions for the posterior distribution on the level set function. Thus, in order to produce samples from the posterior arising in the Bayesian approach, MCMC methods are often used. Since the posterior is typically defined on an infinite-dimensional space in the context of inverse problems, it is important that the MCMC algorithms used are well defined on such spaces. A formulation of the Metropolis–Hastings algorithm on general state spaces is given in Tierney (1998). A particular case of this

algorithm, well suited to posterior distributions on function spaces and Gaussian priors, is the preconditioned Crank–Nicolson (pCN) method introduced (although not named this way) in Beskos et al. (2008). As the method is defined directly on a function space, it has desirable properties related to discretization—in particular the method is robust with respect to mesh refinement (discretization invariance)—see Cotter et al. (2013) and the references therein. On the other hand, the need for hierarchical models in Bayesian statistics, and in particular in the context of non-parametric (i.e., function space) methods in machine learning, is well established Bishop (2006). However, care is needed when using hierarchical methods in order to ensure that discretization invariance is not lost Agapiou et al. (2014). In this paper we demonstrate how hierarchical methods can be employed in the context of discretization-invariant MCMC methods for Bayesian level set inversion.

1.2 Key contributions of the paper

The key contribution of this paper is in computational statistics: we develop a Metropolis Hastings method with mesh-independent mixing properties that makes an order of magnitude of improvement in the Bayesian level set method as introduced in Iglesias et al. (2016).

Study of Fig. 1 suggests that the ability of the level set representation to accurately reconstruct piecewise continuous fields depends on two important scale parameters:

- the length scale of the level set function, and its relation to the typical separation between discontinuities;
- the amplitude scale of the level set function, and its relation to the levels used for thresholding.

If these two scale parameters are not set correctly, then MCMC methods to determine the level set function from data can perform poorly. This immediately suggests the idea of using hierarchical Bayesian methods in which these parameters are learned from the data. However, there is a second consideration which interacts with this discussion. From the work of Tierney [Tierney \(1998\)](#) it is known that absolute continuity of certain measures arising in the definition of Metropolis–Hastings methods is central to their well-definedness, and hence to discretization-invariant MCMC methods [Cotter et al. \(2013\)](#). In fact it appears algorithms defined on infinite-dimensional spaces have spectral gaps that are bounded independently of the mesh, and so their convergence rates are bounded below in the limit [Hairer et al. \(2014\)](#). The key contribution of our paper is to show how enforcing absolute continuity links the two scale parameters, and hence leads to the construction of a hierarchical Bayesian level set method with a *single scalar* hierarchical parameter which deals with the scale and absolute continuity issues simultaneously, resulting in effective sampling algorithms.

The hierarchical parameter is an inverse length scale within a Gaussian random field prior for the level set function. In order to preserve absolute continuity of different priors on the level set function as the length-scale parameter varies, and relatedly to make well-defined MCMC methods, the mean square amplitude of this Gaussian random field must decay proportionally to a power of the inverse length scale. It is thus natural that the level values used for thresholding should obey this power law relationship with respect to the hierarchical parameter. As a consequence the likelihood depends on the hierarchical parameter, leading to a novel form of posterior distribution.

We construct this posterior distribution and demonstrate how to sample from it using a Metropolis-within-Gibbs algorithm which alternates between updating the level set function and the inverse length scale. As a second contribution of the paper, we demonstrate the applicability of the algorithm on three inverse problems, by means of simulation studies. The first concerns reconstruction of a ternary piecewise constant field from a finite noisy set of point measurements: in this context, the Bayesian level set method is very closely related to a spatial probit model [Rasmussen and Williams \(2006\)](#). This relation is discussed in Sect. 2.4. The other two concern reconstruction of the coefficient of a divergence form elliptic PDE from measurements of its solution; in particular, groundwater flow (in which measurements are made in the interior of the

domain) and EIT (in which measurements are made on the boundary).

1.3 Structure of the paper

In Sect. 2 we describe a family of prior distributions on the level set function, indexed by an inverse length-scale parameter, which remain absolutely continuous with respect to one another when we vary this parameter; we then place a hyperprior on this parameter. We describe an appropriate level set map, dependent on the length-scale parameter because length and amplitude scales are intimately connected through absolute continuity of measures, to transform these fields into piecewise constant ones, and use this level set map in the construction of the likelihood. We end by showing existence and well-posedness of the posterior distribution on the level set function and the inverse length-scale parameter. In Sect. 3 we describe a Metropolis-within-Gibbs MCMC algorithm for sampling the posterior distribution, taking advantage of existing state-of-the-art function space MCMC, and the absolute continuity of our prior distributions with respect to changes in the inverse length-scale parameter, established in the previous section. Section 4 contains numerical experiments for three different forward models: a linear map comprising pointwise observations, groundwater flow, and EIT; these illustrate the behavior of the algorithm and, in particular, demonstrate significant improvement with respect to non-hierarchical Bayesian level set inversion.

2 Construction of the posterior

In Sect. 2.1 we recall the definition of the Whittle–Matérn covariance functions, and define a related family of covariances parametrized by an inverse length-scale parameter τ . We use these covariances to define our prior on the level set function u , and also place a hyperprior on the parameter τ , yielding a prior $\mathbb{P}(u, \tau)$ on a product space. In Sect. 2.2 we construct the level set map, taking into account the amplitude scaling of prior samples with τ , and incorporate this into the forward map. The inverse problem is formulated, and the resulting likelihood $\mathbb{P}(y|u, \tau)$ is defined. Finally in Sect. 2.3 we construct the posterior $\mathbb{P}(u, \tau|y)$ by combining the prior $\mathbb{P}(u, \tau)$ and likelihood $\mathbb{P}(y|u, \tau)$ using Bayes' formula. Well-posedness of this posterior is established.

2.1 Prior

As discussed in the introduction it can be important, within the context of Bayesian level set inversion, to attempt to learn the length scale of the level set function whose level

sets determine interfaces in piecewise continuous reconstructions. This is because we typically do not know a priori the typical separation of interfaces. It is also computationally expedient to work with Gaussian random field priors for the level set function, as demonstrated in Iglesias et al. (2016), Dunlop and Stuart (2015). A family of covariances parameterized by length scale is hence required.

A widely used family of distributions, allowing for control over sample regularity, amplitude, and length scale, are Whittle-Matérn distributions. These are a family of stationary Gaussian distributions with covariance function

$$c_{\sigma,v,\ell}(x,y) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{|x-y|}{\ell} \right)^\nu K_\nu \left(\frac{|x-y|}{\ell} \right),$$

where K_ν is the modified Bessel function of the second kind of order ν Matérn (2013), Stein (2012). These covariances interpolate between exponential covariance, for $\nu = 1/2$, and Gaussian covariance, for $\nu \rightarrow \infty$. As a consequence, the regularity of samples increases as the parameter ν increases. The parameter $\ell > 0$ acts as a characteristic length scale (sometimes referred to as the spatial range) and σ as an amplitude scale (σ^2 is sometimes referred to as the marginal variance). On \mathbb{R}^d , samples from a Gaussian distribution with covariance function $c_{\sigma,v,\ell}$ correspond to the solution of a particular stochastic partial differential equation (SPDE). This SPDE can be derived using the Fourier transform and the spectral representation of covariance functions—the paper Lasanen et al. (2014) derives the appropriate SPDE for the covariance function above:

$$\frac{1}{\sqrt{\beta\ell^d}} (I - \ell^2 \Delta)^{(\nu+d/2)/2} v = W, \quad (1)$$

where W is a white noise on \mathbb{R}^d , and

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

Computationally, implementation of this SPDE approach requires restriction to a bounded subset $D \subseteq \mathbb{R}^d$, and hence the provision of boundary conditions for the SPDE in order to obtain a unique solution. Choice of these boundary conditions may significantly affect the autocorrelations near the boundary. The effects for different boundary conditions are discussed in Lasanen et al. (2014). Nonetheless, the computational expediency of the SPDE formulation makes the approach very attractive for applications and, if necessary, boundary effects can be ameliorated by generating the random fields on larger domains which are a superset of the domain of interest.

From (1) it can be seen that the covariance operator corresponding to the covariance function $c_{\sigma,v,\ell}$ is given by

$$\mathcal{D}_{\sigma,v,\ell} = \beta \ell^d (I - \ell^2 \Delta)^{-\nu-d/2}. \quad (2)$$

The fact that the scalar multiplier in front of the covariance operator $\mathcal{D}_{\sigma,v,\ell}$ changes with the length scale means that the family of measures $\{N(0, \mathcal{D}_{\sigma,v,\ell})\}_\ell$, for fixed σ and ν , are mutually singular. This leads to problems when trying to design hierarchical methods based around these priors. We hence work instead with the modified covariances

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

where $\tau = 1/\ell > 0$ now represents an *inverse* length scale, and $\alpha = \nu + d/2$ still controls the sample regularity. To be concrete we will always assume that the domain of the Laplacian is chosen so that $\mathcal{C}_{\alpha,\tau}$ is well defined for all $\tau \geq 0$; for example, we may choose a periodic box, with domain restricted to functions which integrate to zero over the box, Neumann boundary conditions on a box, again with domain restricted to functions which integrate to zero over the box, or Dirichlet boundary conditions. We have the following theorem concerning the family of Gaussians $\{N(0, \mathcal{C}_{\alpha,\tau})\}_{\tau \geq 0}$, proved in Appendix.

Theorem 1 *Let $D = \mathbb{T}^d$ be the d -dimensional torus, and fix $\alpha > 0$. Define the family of Gaussian measures $\mu_0^\tau = N(0, \mathcal{C}_{\alpha,\tau})$, $\tau \geq 0$. Then*

- (i) *for $d \leq 3$, the $\{\mu_0^\tau\}_{\tau \geq 0}$ are mutually equivalent;*
- (ii) *if $u \sim \mu_0^\tau$, then μ_0^τ -a.s. we have $u \in H^s(D)$ and $u \in C^{[s], s-[s]}(D)$ for all $s < \alpha - d/2$.¹*
- (iii) *if $u \sim \mu_0^\tau$ and $v \sim N(0, \mathcal{D}_{\sigma,v,\ell})$, then*

$$\mathbb{E}\|u\|^2 \propto \tau^{d-2\alpha} \cdot \mathbb{E}\|v\|^2$$

with constant of proportionality independent of τ .

Remark 1 (a) Proof of this theorem is driven by the smoothness of the eigenfunctions of the Laplacian subject to periodic boundary conditions, together with the growth of the eigenvalues, which is like $j^{2/d}$. These properties extend to Laplacians on more general domains and with more general boundary conditions, and to Laplacians with lower order perturbations, and so the above result still holds in these cases. For discussion of this in relation to (ii) see Dashti and Stuart (2016); for parts (i) and (iii) the reader can readily extend the proof given in Appendix.

- (b) The proportionality in part (iii) above could be simplified if it were the case that $\mathbb{E}\|v\|^2$ were independent of τ . However, since we restrict to a bounded domain

¹ i.e., the function has s weak (possibly fractional) derivatives in the Sobolev sense, and the $[s]$ th classical derivative is Hölder with exponent $s - [s]$.

$D \subset \mathbb{R}^d$, boundary effects mean that this is not necessarily true. Neumann boundary conditions, for example, inflate the variance up to a distance of approximately $\ell\sqrt{8\nu} = \sqrt{8\nu}/\tau$ from the boundary Lindgren and Rue (2015). Nonetheless, at points $x \in D$ sufficiently far away from the boundary we have $\mathbb{E}|v(x)|^2 \approx \sigma^2$ independently of x . At these points we would hence expect that, for $u \sim \mu_0^\tau$,

$$\mathbb{E}|u(x)|^2 \propto \tau^{d-2\alpha}.$$

Note also that numerically, we may produce samples on a larger domain D^* that contains the domain of interest D , in order to minimize the boundary effects within D .

Let $X = C(D)$ denote the space of continuous real-valued functions on domain D . In what follows we will always assume that $\alpha - d/2 > 0$ in order that the measures have samples in X almost surely. Additionally we shall write \mathcal{C}_τ in place of $\mathcal{C}_{\alpha,\tau}$ when the parameter α is not of interest.

In Sect. 2.2, we pass the inverse length-scale parameter τ to the forward map and treat it as an additional unknown in the inverse problem. We therefore require a joint prior $\mathbb{P}(u, \tau)$ on both the level set field and on τ . We will treat τ as a hyper-parameter, so that $\mathbb{P}(u, \tau)$ takes the form $\mathbb{P}(u, \tau) = \mathbb{P}(u|\tau)\mathbb{P}(\tau)$. Specifically, we will take the conditional distribution $\mathbb{P}(u|\tau)$ to be given by $\mu_0^\tau = N(0, \mathcal{C}_\tau)$, and the hyperprior $\mathbb{P}(\tau)$ to be any probability measure π_0 on \mathbb{R}^+ , the set of positive reals; in practice it will always have a Lebesgue density on \mathbb{R}^+ . The joint prior μ_0 on $X \times \mathbb{R}^+$ is therefore assumed to be given by

$$\mu_0(du, d\tau) = \mu_0^\tau(du)\pi_0(d\tau). \tag{3}$$

Non-zero means could also be considered via a change of coordinates. Discussion of prior choice for the hierarchical parameters in latent Gaussian models may be found in Fuglstad et al. (2015).

2.2 Likelihood

In the previous subsection we defined a prior distribution μ_0 on $X \times \mathbb{R}^+$. We now define a way of constructing a piecewise constant field from a sample (u, τ) . In Iglesias et al. (2016), where the Bayesian level set method was introduced, the piecewise constant field was constructed purely as a function of u as follows. Let $n \in \mathbb{N}$ and fix constants $-\infty = c_0 < c_1 < \dots < c_n = \infty$. Given $u \in X$, define $D_i(u) \subseteq D$ by

$$D_i(u) = \{x \in D \mid c_{i-1} \leq u(x) < c_i\}, \quad i = 1, \dots, n$$

so that $\bar{D} = \bigcup_{i=1}^n \bar{D}_i(u)$ and $D_i(u) \cap D_j(u) = \emptyset$ for $i \neq j, i, j \geq 1$. Then given $\kappa_1, \dots, \kappa_n \in \mathbb{R}$, define the map $F : X \rightarrow Z$ by

$$F(u) = \sum_{i=1}^n \kappa_i \mathbb{1}_{D_i(u)}. \tag{4}$$

Thus F maps the level set field to the geometric field, which is the field of interest, even though inference is performed on the level set field. We may take $Z = L^p(D)$, the space of p -integrable functions on D , for any $1 \leq p \leq \infty$. $F(u)$ then defines a piecewise constant function on D ; the interfaces defined by the jumps are given by the level sets $\{x \in D \mid u(x) = c_i\}$.

Remark 2 One of the constraints of this construction, discussed in Iglesias et al. (2016), is that in order for $F(u)$ to pass from κ_i to κ_j , it must pass through all of $\kappa_{i+1}, \dots, \kappa_{j-1}$ first. Thus this construction cannot represent, for example, a triple junction. This also means that that it must be known a priori that, for example, level i is typically found near levels $i - 1$ and $i + 1$, but unlikely to be found near levels $i + 3$ or $i + 4$. This is potentially a significant constraint; we discuss how this may be dealt with in the conclusions.

This construction is effective for a fixed value of τ , but in light of Theorem 1(iii), the amplitude of samples from $N(0, \mathcal{C}_{\alpha,\tau})$ varies with τ . More specifically, since $d - 2\alpha < 0$ by assumption, samples will decay towards zero as τ increases. For this reason, employing fixed levels $\{c_i\}_{i=0}^n$ and then changing the value of τ during a sampling method may render the levels out of reach. We can compensate for this by allowing the levels to change with τ , so that they decay towards zero at the same rate as the samples.

From Theorem 1(iii) and Remark 1(b) we deduce that samples u from $N(0, \mathcal{C}_{\alpha,\tau})$ decay towards zero at a rate of approximately $\tau^{d/2-\alpha}$ with respect to τ . This suggests allowing for the following dependence of the levels on the length-scale parameter τ :

$$c_i(\tau) = \tau^{d/2-\alpha} c_i, \quad i = 1, \dots, n. \tag{5}$$

In order to update these levels, we must pass the parameter τ to the level set map F . We therefore redefine the level set map $F : X \times \mathbb{R}^+ \rightarrow Z$ as follows. Let $n \in \mathbb{N}$, fix initial levels $-\infty = c_0 < c_1 < \dots < c_n = \infty$ and define $c_i(\tau)$ by (5) for $\tau > 0$. Given $u \in X$ and $\tau > 0$, define $D_i(u, \tau) \subseteq D$ by

$$D_i(u, \tau) = \{x \in D \mid c_{i-1}(\tau) \leq u(x) < c_i(\tau)\}, \quad i = 1, \dots, n, \tag{6}$$

² For any subset $A \subset \mathbb{R}^d$ we will denote by \bar{A} its closure in \mathbb{R}^d .

so that $\overline{D} = \bigcup_{i=1}^n \overline{D}_i(u, \tau)$ and $D_i(u, \tau) \cap D_j(u, \tau) = \emptyset$ for $i \neq j, i, j \geq 1$. Now given $\kappa_1, \dots, \kappa_n \in \mathbb{R}$, we define the map $F : X \times \mathbb{R}^+ \rightarrow Z$ by

$$F(u, \tau) = \sum_{i=1}^n \kappa_i \mathbb{1}_{D_i(u, \tau)}. \tag{7}$$

We can now define the likelihood. Let $Y = \mathbb{R}^J$ be the data space, and let $S : Z \rightarrow Y$ be a forward operator. Define $\mathcal{G} : X \times \mathbb{R}^+ \rightarrow Y$ by $\mathcal{G} = S \circ F$. Assume we have data $y \in Y$ arising from observations of some $(u, \tau) \in X \times \mathbb{R}^+$ under \mathcal{G} , corrupted by Gaussian noise $\eta \sim \mathbb{Q}_0 := N(0, \Gamma)$ on Y :

$$y = \mathcal{G}(u, \tau) + \eta. \tag{8}$$

We now construct the likelihood $\mathbb{P}(y|u, \tau)$. In the Bayesian formulation, we place a prior μ_0 of the form (3) on the pair (u, τ) . Assuming \mathbb{Q}_0 is independent of μ_0 , the conditional distribution $\mathbb{Q}_{u, \tau}$ of y given (u, τ) is given by

$$\frac{d\mathbb{Q}_{u, \tau}}{d\mathbb{Q}_0}(y) = \exp\left(-\Phi(u, \tau; y) + \frac{1}{2}|y|_{\Gamma}^2\right), \tag{9}$$

where the potential (or negative log-likelihood) $\Phi : X \times \mathbb{R}^+ \rightarrow \mathbb{R}$ is defined by

$$\Phi(u, \tau; y) = \frac{1}{2}|y - \mathcal{G}(u, \tau)|_{\Gamma}^2 \tag{10}$$

and $|\cdot|_{\Gamma} := |\Gamma^{-1/2} \cdot|$.

Denote $\text{Im}(F) \subseteq Z$ the image of $F : X \times \mathbb{R}^+ \rightarrow Z$. In what follows we make the following assumptions on $S : Z \rightarrow Y$.

- Assumption 1** (i) S is continuous on $\text{Im}(F)$.
(ii) For any $r > 0$ there exists $C(r) > 0$ such that for any $z \in \text{Im}(F)$ with $\|z\|_{L^\infty} \leq r, |S(z)| \leq C(r)$.

In the next subsection we show that, under the above assumptions, the posterior distribution μ^y of (u, τ) given y exists, and study its properties.

2.3 Posterior

Bayes' theorem provides a way to construct the posterior distribution $\mathbb{P}(u, \tau|y)$ using the ingredients of the prior $\mathbb{P}(u, \tau)$ and the likelihood $\mathbb{P}(y|u, \tau)$ from the previous two subsections. Informally we have

$$\begin{aligned} \mathbb{P}(u, \tau|y) &\propto \mathbb{P}(y|u, \tau)\mathbb{P}(u, \tau) \\ &\propto \exp(-\Phi(u, \tau; y))\mu_0^\tau(u)\pi_0(\tau), \end{aligned}$$

after absorbing y -dependent constants from the likelihood into the normalization constant. In order to make this formula rigorous some care must be taken, since μ_0^τ does not admit a Lebesgue density. The following is proved in Appendix.

Theorem 2 *Let μ_0 be given by (3), y by (8) and Φ be given by (10). Let Assumptions 1 hold. If $\mu^y(du, d\tau)$ is the regular conditional probability measure on $(u, \tau)|y$, then $\mu^y \ll \mu_0$ with Radon–Nikodym derivative*

$$\frac{d\mu^y}{d\mu_0}(u, \tau) = \frac{1}{Z} \exp(-\Phi(u, \tau; y)),$$

where, for y almost surely,

$$Z := \int_{X \times \mathbb{R}^+} \exp(-\Phi(u, \tau; y)) \mu_0(du, d\tau) > 0.$$

Furthermore μ^y is locally Lipschitz with respect to y , in the Hellinger distance: for all y, y' with $\max\{|y|_{\Gamma}, |y'|_{\Gamma}\} < r$, there exists a $C = C(r) > 0$ such that

$$d_{\text{Hell}}(\mu^y, \mu^{y'}) \leq C|y - y'|_{\Gamma}.$$

This implies that, for all $f \in L^2_{\mu_0}(X \times \mathbb{R}^+; E)$ for separable Banach space E ,

$$\|\mathbb{E}^{\mu^y} f(u, \tau) - \mathbb{E}^{\mu^{y'}} f(u, \tau)\|_E \leq C|y - y'|.$$

To the best of our knowledge this form of Bayesian posterior distribution, in which the prior hyper-parameter appears in the likelihood because it is natural to scale a thresholding function with that parameter, for algorithmic reasons, is novel. A different form of thresholding is studied in the paper Bolin and Lindgren (2015) where boundaries defining regions in which certain events occur with a specified (typically close to 1) probability is studied.

2.4 Relation to probit models

The Bayesian level set method has a close relation with an ordered probit model in the case that the state space X is finite dimensional. Suppose that $X = \mathbb{R}^N$, then neglecting the length-scale parameter, the data y_{level} in the level set method are assumed to arise via

$$y_{\text{level}} = \mathcal{G}(F(u)) + \eta, \quad \eta \sim N(0, \Gamma),$$

where F denotes the original thresholding function as defined by (4). In an ordered probit model, the data y_{probit} are assumed to arise via³

³ The thresholding function F is defined pointwise, so can be considered to be defined on either \mathbb{R}^N or \mathbb{R} , with $F(u)_n = F(u_n)$.

$$y_{\text{prob}} = \mathcal{G}(z),$$

$$z_n = F(u_n + \varepsilon_n), \quad \varepsilon_n \sim N(0, 1), \quad n = 1, \dots, N.$$

Note that in the case of probit, the noise is applied before the thresholding F so that the geometric field takes values in the discrete set $\{\kappa_1, \dots, \kappa_n\}$. In contrast in the case of the level set model the noise is applied after thresholding. If \mathcal{G} is linear then the probit model results in categorical data, while in the level set case the data can take any real value. Depending on the forward model either probit or level set may be more appropriate: the former in cases where the data are genuinely discrete and interpolation between phases does not have a meaning, such as categorical data, and the latter when they are continuous, such as when corrupted by measurement noise. The two models could also be combined, which may be interesting in some applications. In the small noise limit the models are seen to be equivalent.

Placing a prior upon u leads to a well-defined posterior distribution in both cases. Dimension-robust sampling of both distributions can be performed using a prior-reversible MCMC method, such as the preconditioned Crank–Nicolson (pCN) method [Cotter et al. \(2013\)](#). The spatial version of probit, that is when X is a function space rather than \mathbb{R}^N , is of interest to study further.

Once we introduce the hierarchical length-scale dependence, significant problems arise in terms of sampling the probit posterior in high dimensions, due to the issues associated with measure singularity discussed above. With the level set method it is possible to circumvent through the choice of prior and rescaling discussed in this section; a well-defined Metropolis-within-Gibbs sampling algorithm on function space is outlined in the next section.

3 MCMC algorithm for posterior sampling

Having constructed the posterior distribution on $(u, \tau)|y$, we are now faced with the task of sampling this probability distribution. We will use the Metropolis-within-Gibbs formalism, as described in for example [Robert and Casella \(2013\)](#), Sect. 10.3. This algorithm constructs the Markov chain $(u^{(k)}, \tau^{(k)})$ with the structure

$$\begin{aligned} - u^{(k+1)} &\sim \mathbb{K}^{\tau^{(k)}, y}(u^{(k)}, \cdot), \\ - \tau^{(k+1)} &\sim \mathbb{L}^{u^{(k+1)}, y}(\tau^{(k)}, \cdot), \end{aligned}$$

where $\mathbb{K}^{\tau, y}$ is a Metropolis–Hastings Markov kernel reversible with respect to $u|\tau, y$ and $\mathbb{L}^{u, y}$ is a Metropolis–Hastings Markov kernel reversible with respect to $\tau|u, y$. The Metropolis–Hastings method is outlined in chapter 7 of [Robert and Casella \(2013\)](#). See [Geirsson et al. \(2015\)](#) for

related blocking methodologies for Gibbs samplers in the context of latent Gaussian models.

In defining the conditional distributions, and the Metropolis methods to sample from them, a key design principle is to ensure that all measures and algorithms are well defined in the infinite-dimensional setting, so that the resulting algorithms are robust to mesh refinement [Cotter et al. \(2013\)](#). This thinking has been behind the form of the prior and posterior distributions developed in the previous section, as we now demonstrate.

In Sect. 3.1 we define the kernel $\mathbb{K}^{\tau, y}$ and in Sect. 3.2 we define the kernel $\mathbb{L}^{u, y}$. Then in the final Sect. 3.3 we put all these building blocks together to specify the complete algorithm used.

3.1 Proposal and acceptance probability for $u|\tau, y$

Samples from the distribution of $u|\tau, y$ can be produced using a pCN Metropolis Hastings method [Cotter et al. \(2013\)](#), with proposal and acceptance probability as follows:

1. Given u , propose

$$v = (1 - \beta^2)^{1/2}u + \beta\xi, \quad \xi \sim N(0, \mathcal{C}_\tau).$$

2. Accept with probability

$$\alpha(u, v) = \min\{1, \exp(\Phi(u, \tau; y) - \Phi(v, \tau; y))\}$$

or else stay at u .

3.2 Proposal and acceptance probability for $\tau|y, u$

Producing samples of $\tau|u, y$ is more involved, since we must first make sense of this conditional distribution. To do this, define the three measures η_0, ν_0 , and ν on $X \times \mathbb{R}^+ \times Y$ by

$$\begin{aligned} \eta_0(du, d\tau, dy) &= \mu_0^0(du)\pi_0(d\tau)\mathbb{Q}_0(dy), \\ \nu_0(du, d\tau, dy) &= \mu_0^\tau(du)\pi_0(d\tau)\mathbb{Q}_0(dy), \\ \nu(du, d\tau, dy) &= \mu_0^\tau(du)\pi_0(d\tau)\mathbb{Q}_{u, \tau}(dy). \end{aligned}$$

Here $\mathbb{Q}_0 = N(0, \Gamma)$ is the distribution of the noise, and $\mathbb{Q}_{u, \tau}$ is as defined in (9). Then we have the chain of absolute continuities $\nu \ll \nu_0 \ll \eta_0$, with

$$\frac{d\nu_0}{d\eta_0}(u, \tau, y) = \frac{d\mu_0^\tau}{d\mu_0^0}(u) =: L(u, \tau),$$

$$\frac{d\nu}{d\nu_0}(u, \tau, y) = \frac{d\mathbb{Q}_{u, \tau}}{d\mathbb{Q}_0}(y) = \exp\left(-\Phi(u, \tau; y) + \frac{1}{2}|y|_\Gamma^2\right),$$

and so by the chain rule we have $\nu \ll \eta_0$ and

$$\frac{d\nu}{d\eta_0}(u, \tau, y) = \frac{d\mathbb{Q}_{u, \tau}}{d\mathbb{Q}_0}(y) \cdot \frac{d\mu_0^\tau}{d\mu_0^0}(u) =: \varphi(u, \tau, y).$$

We use the conditioning lemma, Theorem 3.1 in [Dashti and Stuart \(2016\)](#), to prove the existence of the desired conditional distribution.

Theorem 3 Assume that $\Phi : X \times \mathbb{R}^+ \times Y \rightarrow \mathbb{R}$ is η_0 measurable and η_0 -a.s. finite. Assume also that, for $(u, y) \mu_0^0 \times \mathbb{Q}_0$ -a.s.,

$$Z_\pi := \int_{\mathbb{R}^+} \exp(-\Phi(u, \tau; y)) L(u, \tau) \pi_0(d\tau) > 0.$$

Then the regular conditional distribution of $\tau|(u, y)$ exists under ν , and is denoted by $\pi^{u,y}$. Furthermore, $\pi^{u,y} \ll \pi_0$ and, for $(u, y) \nu$ -a.s.,

$$\frac{d\pi^{u,y}}{d\pi_0}(\tau) = \frac{1}{Z_\pi} \exp(-\Phi(u, \tau; y)) L(u, \tau).$$

Proof The conditional random variable $\tau|(u, y)$ exists under η_0 , and its distribution is just π_0 since η_0 is a product measure. Theorem 3.1 in [Dashti and Stuart \(2016\)](#) then tells us that the conditional random variable $\tau|(u, y)$ exists under ν . We denote its distribution $\pi^{u,y}$. Define

$$\begin{aligned} c(u, y) &= \int_{\mathbb{R}^+} \varphi(u, \tau, y) \pi_0(d\tau) \\ &= \exp\left(\frac{1}{2}|y|_T^2\right) \int_{\mathbb{R}^+} \exp(-\Phi(u, \tau; y)) L(u, \tau) \pi_0(d\tau). \end{aligned}$$

Now since $\exp(\frac{1}{2}|y|_T^2) \in (0, \infty) \mu_0^0 \times \mathbb{Q}_0$ -a.s., we deduce that $c(u, y) > 0 \mu_0^0 \times \mathbb{Q}_0$ -a.s. by the $\mu_0^0 \times \mathbb{Q}_0$ -a.s. positivity of Z_π . By the absolute continuity $\nu \ll \eta_0$, we deduce that $c(u, y) > 0 \nu$ -a.s. Therefore, again by Theorem 3.1 in [Dashti and Stuart \(2016\)](#), we have $\pi^{u,y} \ll \pi_0$ and, for $(u, y) \nu$ -a.s.,

$$\begin{aligned} \frac{d\pi^{u,y}}{d\pi_0}(\tau) &= \frac{1}{c(u, y)} \varphi(u, \tau, y) \\ &= \frac{1}{Z_\pi} \exp(-\Phi(u, \tau; y)) L(u, \tau). \end{aligned}$$

□

Remark 3 Above we have used μ_0^0 as a reference measure, and the function $L(u, \tau)$ enters our expression for the posterior. But any μ_0^λ will suffice since the entire family of measures $\{\mu_0^\tau\}_{\tau \geq 0}$ are equivalent to one another. A straightforward calculation with the chain rule gives

$$\begin{aligned} \frac{d\pi^{u,y}}{d\pi_0}(\tau) &= \frac{1}{Z_{\pi,\lambda}} \frac{d\mu_0^\tau}{d\mu_0^\lambda}(u) \exp(-\Phi(u, \tau; y)) \\ &= \frac{1}{Z_{\pi,\lambda}} L_\lambda(u, \tau) \exp(-\Phi(u, \tau; y)). \end{aligned}$$

We now wish to sample from $\pi^{u,y}$ using a Metropolis–Hastings algorithm. We assume from now on that π_0 admits

a Lebesgue density, so that $\pi^{u,y}$ also admits a Lebesgue density. Abusing notation and using $\pi^{u,y}$, π_0 to denote Lebesgue densities as well as the corresponding measures we have

$$\pi^{u,y}(\tau) \propto \exp(-\Phi(u, \tau; y)) L(u, \tau) \pi_0(\tau).$$

Take a proposal kernel $Q(\tau, d\gamma) = q(\tau, \gamma) d\gamma$. Define the two measures ρ, ρ^T on $(\mathbb{R} \times \mathbb{R}, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}))$ by

$$\begin{aligned} \rho(d\tau, d\gamma) &= \pi^{u,y}(d\tau) Q(\tau, d\gamma) \\ &\propto \exp(-\Phi(u, \tau; y)) L(u, \tau) \pi_0(\tau) q(\tau, \gamma) d\tau d\gamma, \\ \rho^T(d\tau, d\gamma) &= \mu(d\gamma, d\tau). \end{aligned}$$

Then under appropriate conditions on π_0 and q , these two measures are equivalent. Define $r(\tau, \gamma)$ to be the Radon–Nikodym derivative

$$\begin{aligned} r(\tau, \gamma) &:= \frac{d\rho^T}{d\rho}(\tau, \gamma) \\ &= \exp(\Phi(u, \tau; y) - \Phi(u, \gamma; y)) \times \frac{d\mu_0^\gamma}{d\mu_0^\tau}(u) \\ &\quad \times \frac{\pi_0(\gamma) q(\gamma, \tau)}{\pi_0(\tau) q(\tau, \gamma)}. \end{aligned}$$

The general form of the Metropolis–Hastings algorithm, as for example given in [Tierney \(1998\)](#), says that we produce samples from $\pi^{u,y}$ by iterating the follow two steps:

1. Given τ , propose $\gamma \sim Q(\tau, d\gamma)$.
2. Accept with probability $\alpha(\tau, \gamma) = n\{1, r(\tau, \gamma)\}$, or else stay at τ .

In order to implement this algorithm, we need an expression for the Radon–Nikodym derivative $\frac{d\mu_0^\gamma}{d\mu_0^\tau}(u)$. Denote by $\{\lambda_j(\tau)\}_{j \geq 1}$ the eigenvalues of the covariance \mathcal{C}_τ , and $\{\varphi_j\}_{j \geq 1}$ their corresponding eigenvectors. Note that because of the structure of the family $\{\mathcal{C}_\tau\}_{\tau \geq 0}$, the eigenvectors are independent of τ . Using Proposition 3, we see that

$$\begin{aligned} \frac{d\mu_0^\gamma}{d\mu_0^\tau}(u) &= \prod_{j=1}^{\infty} \frac{\lambda_j(\tau)^{1/2}}{\lambda_j(\gamma)^{1/2}} \\ &\quad \times \exp\left(\frac{1}{2} \sum_{j=1}^{\infty} \left(\frac{1}{\lambda_j(\tau)} - \frac{1}{\lambda_j(\gamma)}\right) \langle u, \varphi_j \rangle^2\right) \\ &= \exp\left(\frac{1}{2} \left[\sum_{j=1}^{\infty} \left(\frac{1}{\lambda_j(\tau)} - \frac{1}{\lambda_j(\gamma)}\right) \langle u, \varphi_j \rangle^2 \right. \right. \\ &\quad \left. \left. + \log\left(\frac{\lambda_j(\tau)}{\lambda_j(\gamma)}\right) \right]\right). \end{aligned} \tag{11}$$

From Theorem 1 we know that μ_0^τ and μ_0^γ are equivalent, and so it must be the case that the expressions for the derivative above are almost-surely finite. However, this is not immediately clear from inspection of the expression; thus we provide some intuition about why it is so in the following theorem. The proof is given in Appendix.

Theorem 4 Assume that $u \sim N(0, C_0)$. Then for each $\tau > 0$,

- (i) $\sum_{j=1}^\infty \left(\frac{1}{\lambda_j(\tau)} - \frac{1}{\lambda_j(0)} \right) \langle u, \varphi_j \rangle^2$ is almost-surely finite if and only if $d = 1$; and
- (ii) $\sum_{j=1}^\infty \left[\left(\frac{1}{\lambda_j(\tau)} - \frac{1}{\lambda_j(0)} \right) \langle u, \varphi_j \rangle^2 + \log \left(\frac{\lambda_j(\tau)}{\lambda_j(0)} \right) \right]$ is almost-surely finite if $d \leq 3$.

A consequence of part (i) of this result is that in dimensions 2 and 3, both the product and the sum in (11) diverge, despite the whole expression being finite. This means that care is required when numerically implementing the Gibbs update of τ .

3.3 The algorithm

Putting the theory above together, we can write down a Metropolis-within-Gibbs algorithm for sampling the posterior distribution. Recall that we assumed the proposal kernel Q admitted a Lebesgue density q : $Q(\tau, d\gamma) = q(\tau, \gamma)d\gamma$.

Let $\{\lambda_j(\tau), \varphi_j\}_{j \geq 1}$ denote the eigenbasis associated with C_τ . Define

$$w(\tau, \gamma) = \exp \left(\frac{1}{2} \sum_{j=1}^\infty \left[\left(\frac{1}{\lambda_j(\tau)} - \frac{1}{\lambda_j(\gamma)} \right) \langle u, \varphi_j \rangle^2 + \log \left(\frac{\lambda_j(\tau)}{\lambda_j(\gamma)} \right) \right] \right)$$

and set

$$\alpha^\tau(u, v) = \min \left\{ 1, \exp \left(\Phi(u, \tau; y) - \Phi(v, \tau; y) \right) \right\},$$

$$\alpha^u(\tau, \gamma) = \min \left\{ 1, \exp \left(\Phi(u, \tau; y) - \Phi(u, \gamma; y) \right) \cdot w(\tau, \gamma) \cdot \frac{\pi_0(\tau)q(\tau, \gamma)}{\pi_0(\gamma)q(\gamma, \tau)} \right\}.$$

Fix jump parameter $\beta \in (0, 1]$, and generate $\{u^{(k)}, \tau^{(k)}\}_{k \geq 0}$ as follows:

Algorithm 1 Metropolis-within-Gibbs

1. Set $k = 0$ and pick initial state $(u^{(0)}, \tau^{(0)}) \in X \times \mathbb{R}^+$.
2. Propose $v^{(k)} = (1 - \beta^2)^{1/2}u^{(k)} + \beta\xi^{(k)}$, where $\xi^{(k)} \sim N(0, C_\tau)$.
3. Set $u^{(k+1)} = v^{(k)}$ with probability $\alpha^{\tau^{(k)}}(u^{(k)}, v^{(k)})$, or else set $u^{(k+1)} = u^{(k)}$.
4. Propose $\gamma^{(k)} \sim Q(\tau^{(k)}, \cdot)$.
5. Set $\tau^{(k+1)} = \gamma^{(k)}$ with probability $\alpha^{u^{(k+1)}}(\tau^{(k)}, \gamma^{(k)})$, or else set $\tau^{(k+1)} = \tau^{(k)}$.
6. $k \rightarrow k + 1$ and return to 2.

Then $\{u^{(k)}, \tau^{(k)}\}_{k \geq 0}$ is a Markov chain which is invariant with respect to $\mu^y(du, d\tau)$.

4 Numerical results

We perform a variety of numerical experiments to illustrate the performance of the hierarchical algorithm described in Sect. 3. We focus on three different forward models. The first is pointwise observations composed with the identity—the simplicity of this model allows us to probe the behavior of the algorithm at low computational cost, and such models are also of interest in applications such as image reconstruction—see, for example, Alvarez and Morel (1994), Sapiro (2006) and the references therein. The other two, groundwater flow and EIT, are physical models which have previously been studied extensively, including study of non-hierarchical Bayesian level set methods Iglesias et al. (2016), Dunlop and Stuart (2015). A review of studies on inverse problems associated with EIT is given in Borcea (2002).

The code used for simulations is available on GitHub at <https://github.com/mattdunlop/bayes-hier/releases/v1.0>.

4.1 Discretization of the problem

There are two spaces that we must discretize in order to implement the algorithm. The first is the state space, where the samples will be generated, and the second is the function space associated with the evaluation of the forward model. We briefly outline how this is done.

Our discretization for the state space relies on the Karhunen–Loève expansion of the prior. Suppose we wish to produce samples from a Gaussian measure $N(0, C)$, where C has associated eigenbasis $\{\lambda_j, \varphi_j\}_{j \in \mathbb{N}}$. Then a sample u from this distribution may be represented as

$$u(x) = \sum_{j=1}^\infty \sqrt{\lambda_j} \xi_j \varphi_j(x), \quad \xi_j \sim N(0, 1) \text{ i.i.d.}$$

We discretize the space by truncating and approximating this basis, so that elements of the space are represented as

$$u^N(x) = \sum_{j=1}^N u_j^N \varphi_j^N(x).$$

The inference is then performed on the random variables $\{u_j^N\}_{j=1}^N$. Additionally, in the cases we consider, the eigenvectors associated with all covariances are given by the Fourier basis and so we may use the Fast Fourier Transform for efficient implementation.

The second discretization occurs in the solution of the differential equations. In the EIT example a finite element discretization is used, in which the functions are approximated by expansion in a finite basis. The coefficients of the expansion of the solution to the PDE in this basis are then solved for numerically. The basis is chosen such that each basis element is locally supported—this ensures that matrices arising in the implementation of the method are sparse.

The groundwater flow model uses a finite difference discretization, in which derivatives are approximated by difference quotients. For example, given a uniform grid $\{x_i, y_j\}_{i,j=1}^N$ with spacing $x_{i+1} - x_i = \delta$, we may approximate

$$\frac{\partial h}{\partial x}(x_i, y_j) \approx \frac{h(x_i + \delta, y_j) - h(x_i - \delta, y_j)}{2\delta}.$$

This leads to an approximate solution to the PDE defined on the grid $\{x_i, y_j\}_{i,j=1}^N$.

Finite element, finite difference and even spectral methods outlined above can all be used for any PDE examples; what we use for illustrative purposes in this paper (EIT with finite element and groundwater flow with finite difference) are just examples of numerous possible forward models and discretization combinations.

4.2 Identity map

The first inverse problem is based on reconstruction of a piecewise constant field from noisy pointwise observations.

4.2.1 The forward model

Let $D = [0, 1]^2$ and define a grid of observation points $\{q_j\}_{j=1}^J \subseteq D$. Let $Z = L^p(D)$ for some $1 \leq p < \infty$ and let $Y = \mathbb{R}^J$. The forward operator $S : Z \rightarrow Y$ is defined by

$$S(\kappa) = (\kappa(q_1), \dots, \kappa(q_J)).$$

We are then interested in finding κ , given the prior information that it is piecewise constant, and taking a number of known prescribed values. Let $\mathcal{G} = S \circ F : X \times \mathbb{R}^+ \rightarrow Y$. We reconstruct (u, τ) and hence $\kappa = F(u, \tau)$. The map S is not continuous, and so Assumptions 1 do not hold. However, Proposition 2 in Appendix shows that the map \mathcal{G} is uniformly bounded, and almost-surely continuous under the priors considered. From this the conclusions of

Proposition 1 in Appendix follow, and it is possible to deduce the conclusions of Theorem 2.

4.2.2 Simulations and results

We study the effect of different length scales, for both hierarchical and non-hierarchical methods, demonstrating the advantages of the former over the latter. To this end we define $\tau_i^\dagger = 5i$, $i = 1, \dots, 10$, and generate 10 different true level set fields $u_i^\dagger \sim \mu_0^{\tau_i^\dagger}$ on a mesh of $2^{10} \times 2^{10}$ points. This leads to 10 sets of data y_i , given by

$$y_i = \mathcal{G}(u_i^\dagger, \tau_i^\dagger) + \eta_i, \quad \eta_i \sim N(0, \Gamma) \text{ i.i.d.},$$

where we take the noise covariance $\Gamma = 0.2^2 \cdot I$ to be white. The level set map F is defined such that there are 3 phases, taking the constant values 1, 3, and 5. The mean relative error on the generated datasets ranges from 6 to 9%.

One of the motivations for developing a hierarchical method is that little knowledge may be known a priori about the length scale associated with the unknown geometric field. We therefore sample from each hierarchical posterior distribution associated with each y_i using a variety of initial values for the length-scale parameter. This allows us to check that, computationally, we can recover a good approximation to the true length scale even if our initial guess is poor. Specifically, for each set of data we run 10 hierarchical MCMC simulations started at the different values of $\tau = \tau_k^\dagger$, giving a total of 100 hierarchical MCMC chains. For all chains we place a relatively flat prior of $N(20, 10^2)$ on τ . On the prior for the level set function u , we take Neumann boundary conditions and fix the smoothness parameter $\alpha = 5$. The thresholding levels in the level set map are chosen such that there is an order one amount of prior mass in all levels—specifically we take $c_1 = -0.1$ and $c_2 = 0.1$.

We also wish to compare how the hierarchical method compares with the non-hierarchical method. We therefore look at the 10 different posterior distributions that arise from each set of data y_i when using each of 10 fixed prior inverse length scales τ_k^\dagger , which gives another 100 MCMC chains.

We perform all sampling on a mesh of $2^7 \times 2^7$ points to avoid an inverse crime, discretizing via the discrete Fourier transform (DFT) and retaining all 2^{14} modes. The observation grid $\{q_j\}_{j=1}^{100}$ is taken to be a uniformly spaced grid of 100 points. We use a Gaussian random walk proposal distribution for the length-scale parameter. We make this choice as it is the canonical starting point for MCMC, and it works in this case. It is possible, however, that something more sophisticated may be beneficial. We produce 5×10^6 samples for each chain, and discard the first 10^6 samples as burn-in when calculating quantities of interest.

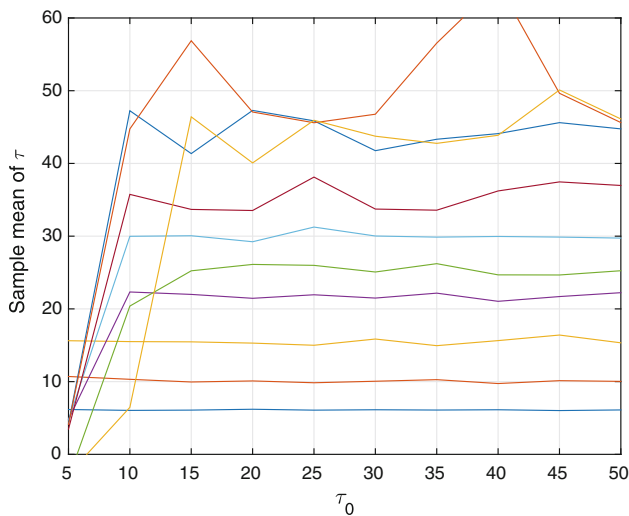


Fig. 2 (Identity model) The sample mean of τ along each hierarchical MCMC chain, against the initial value of τ . The different curves arise from using different data y_i

Table 1 (Identity model) The value of τ used to create the data y_i , and the mean value of τ across the MCMC chains and the different initial values of τ

τ^\dagger	Mean sample mean of τ
5	6.10
10	10.0
15	15.5
20	21.8
25	24.8
30	30.0
35	35.4
40	44.6
45	50.8
50	40.6

In Fig. 2 we look at the recovery of the true value of τ with the hierarchical method. For large enough τ_0 , the mean of τ after the burn-in period is roughly constant with respect to varying the initialization point, for each posterior. This makes sense from a theoretical point of view since these means arise from the same posterior distribution, for a fixed truth, but it is also reassuring from a computational point of view since the output is close to independent of the initial guess for the length scale. There does, however, appear to be an issue with initializing the value of τ at too low a value, with the value τ tending to get stuck far from the truth when initialized at = 5. This effect has been detected in several other experiments and models—initializing the value of τ much lower than the true inverse length can cause the parameter to become stuck in a local minimum. Such an effect has not been observed, however, when the parameter is initialized significantly larger than the true value. Table 1 shows that recovery of the true value of τ is very good for $\tau^\dagger \leq 35$, though becomes slightly

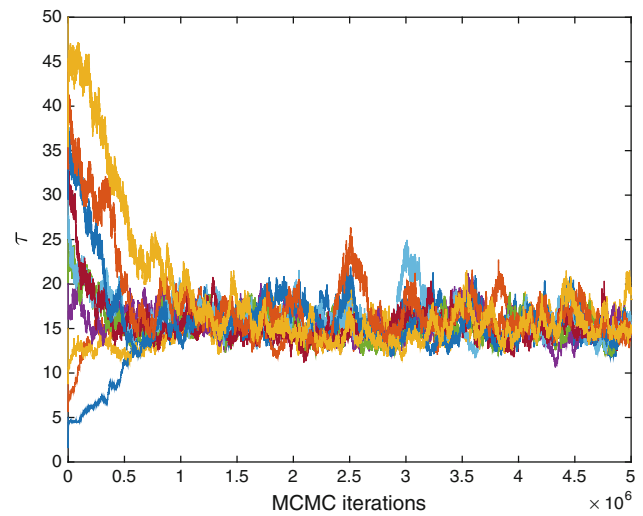


Fig. 3 (Identity model) The trace of τ along the MCMC chain, when initialized at the 10 different initial values. True inverse length scale is $\tau = 15$

worse for larger values of τ^\dagger . The means here are calculated without the $\tau_0 = 5$ sample means since they are clearly outliers for most of the posteriors. One possible explanation for the lack of recovery in the cases $\tau^\dagger = 40, 45$, and 50 is to do with the structure of the observation map S . The observation grid has a length scale associated with it, related to distances between observation points, and so issues could arise when trying to detect the length scale of the geometric field that is significantly shorter than this. Additionally, the length scales $1/\tau$ are closer for larger τ and so it may be more difficult to distinguish between particular values.

For brevity we now focus on the case where $\tau^\dagger = 15$. The traces of the values of τ along the hierarchical chains corresponding to this truth is shown in Fig. 3. After approximately 10^6 samples, all chains have become centered around the true length scale. This convergence appears to be roughly linear for each chain.

Figure 4 shows the push forwards of the sample means from the different chains under the level set map, that is, approximations of $F(\mathbb{E}(u), \mathbb{E}(\tau))$. This figure also shows approximations of $\mathbb{E}(F(u, \tau))$ and typical samples of $F(u, \tau)$ coming from the different chains. We see that these conditional means for the hierarchical method appear to agree with one other. This is reassuring for the reason mentioned above—they are all estimates of the mean of the same distribution. The figures for the non-hierarchical posteriors admit greater variation, especially near the boundary for higher values of τ . Moreover, not all inclusions are detected when the length-scale parameter is taken to be $\tau = 5$. Note that the mean from the hierarchical posterior agrees closely with that from the non-hierarchical posterior using the fixed true length scale $\tau = 15$. Additionally, even though the means are reasonable approximations to the truth in most cases, the typical

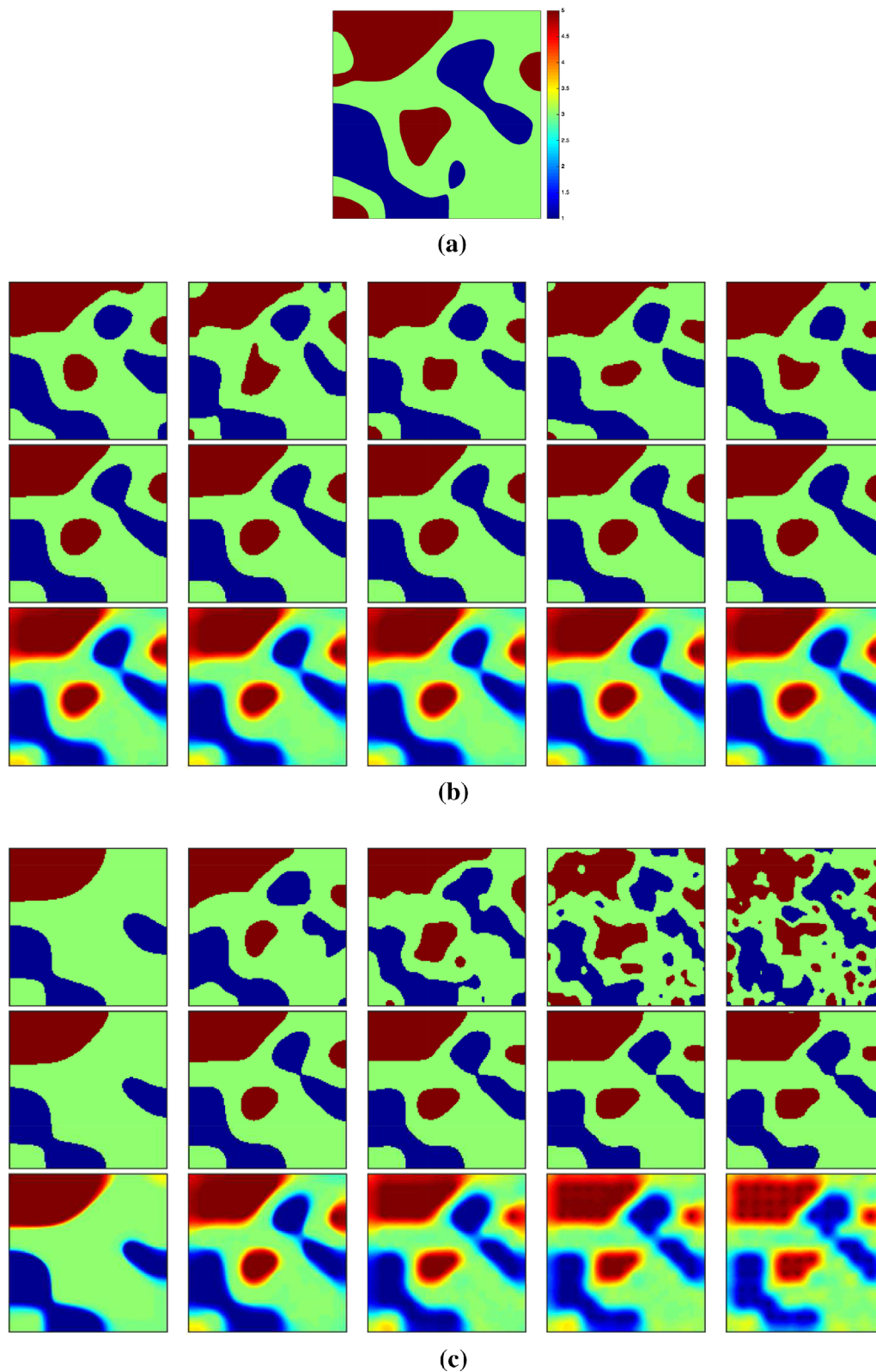


Fig. 4 Simulations for the identity model. **a** The true geometric field used to generate the data y , with true inverse length scale $\tau = 15$. **b** (Top) representative samples of $F(u, \tau)$ under the hierarchical posterior. (Middle) approximations of $F(\mathbb{E}(u), \mathbb{E}(\tau))$. (Bottom) approximations of

$\mathbb{E}(F(u, \tau))$. From left-to-right, τ is initialized at $\tau = 5, 15, 25, 35, 45$. **c** As in **b**, using the non-hierarchical method. From left-to-right, τ is fixed at $\tau = 5, 15, 25, 35, 45$

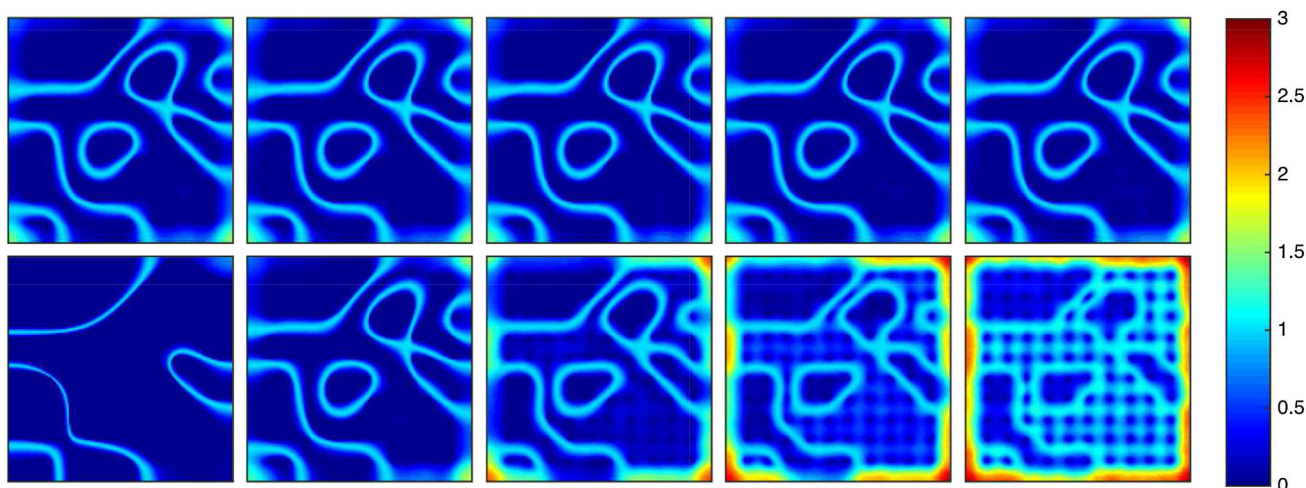


Fig. 5 (Identity model) approximations of $\text{Var}(F(u, \tau))$ using the hierarchical (*top*) and fixed (*bottom*) priors, initialized or fixed at $\tau = 5, 15, 25, 35, 45$, from *left-to-right*. True inverse length scale is $\tau = 15$

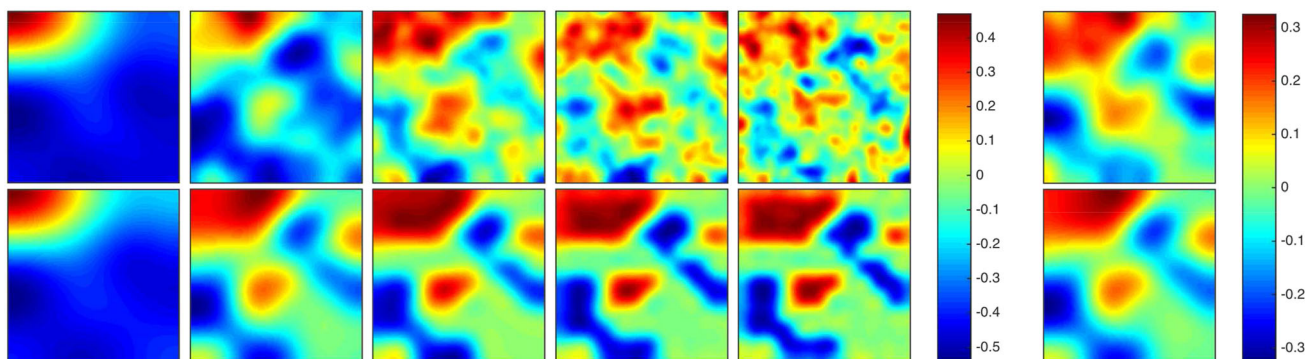


Fig. 6 (Identity model) representative samples $\tau^4 \cdot u$ (*top*) and sample means $\mathbb{E}(\tau^4 \cdot u)$ (*bottom*) of the level set function. The rescaling τ^4 means that the above quantities have the same approximate amplitude.

True inverse length scale is $\tau = 15$. (*Left*) Using the non-hierarchical method; from *left-to-right* τ is fixed at $\tau = 5, 15, 25, 35, 45$. (*Right*) Corresponding quantities for the hierarchical method

samples are much worse when using the non-hierarchical method with an incorrect length-scale parameter.

We can also consider the sample variance of the pushforward of the samples by the level set map, i.e., approximations of the quantity $\text{Var}(F(u, \tau))$. In Fig. 5 we show this quantity for both the hierarchical and non-hierarchical priors. Note that for the non-hierarchical priors, the variance increases both at the boundary and away from the observation points for larger values of τ . Variance is also higher along the interfaces and within the central phase, since points in these locations are more likely to switch between all three phases. The hierarchical approximations all appear to agree. While the hierarchical means are very similar to the non-hierarchical means using the true length scale, as seen in Fig. 4, the hierarchical variances are smaller away from the observation points.

Additionally, we look at the level set function u itself in Fig. 6. In these plots we rescale the level set function by $\tau^{\alpha-d/2} = \tau^4$ so that they are all of approximately the same amplitude. The means for both the hierarchical and non-

hierarchical methods are again quite similar to one another, though the difference between the typical samples is much more stark.

Finally, in Fig. 7, we look at the joint densities of the inverse length-scale parameter τ and first five Karhunen–Loève (KL) modes of the level set function u .⁴ Non-trivial correlations are evident between τ and each of these modes, with the support of the densities appearing non-convex. This is likely related to the non-linear scaling between the length scale and the amplitude of the level set function under the prior. Conversely the KL modes, while still correlated with one another, have simpler joint densities. Note, also, that the posterior on the length scale is centered close to the true value of the inverse length-scale parameter τ .

Remark 4 In this section we studied the ability to recover the true length-scale parameter τ^\dagger , given a finite number of

⁴ KL modes are the eigenfunctions of the covariance operator, here ordered by decreasing eigenvalue.

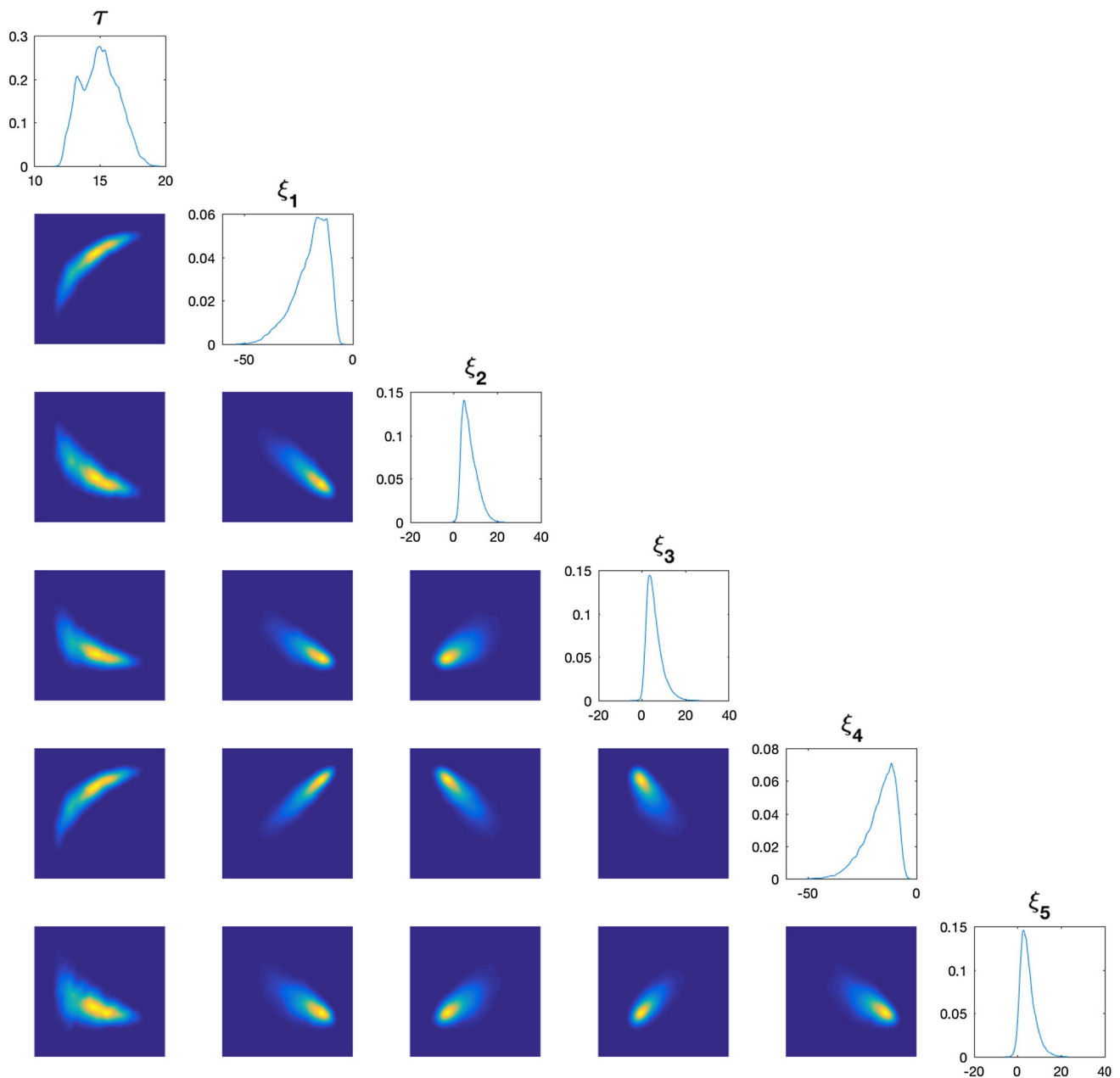


Fig. 7 (Identity model) (diagonal) empirical densities of τ and the first five KL modes of u . (off-diagonal) empirical joint densities. True inverse length scale is $\tau = 15$

direct noisy observations of the geometric field. The question arises of how the quality of this recovery depends upon the spatial resolution of the data. As would be expected, learning this parameter becomes more difficult when this resolution is poor due to the lack of information in the data. However, it is interesting to note that, even in the limit of an infinite number of distinct observation points, it is unlikely that we would be able to identify τ^\dagger perfectly. This is suggested by a result of Zhang (2004) which states that, in the context of generalized linear mixed models, the marginal variance and length-scale parameters of a Matérn field cannot be consistently estimated

in this limit where as in our case the domain is fixed. This is in contrast to the case of additional data points increasing the domain, where consistent estimation is possible Marshall and Mardia (1984).

4.3 Identification of geologic facies in groundwater flow

The identification of geologic facies in subsurface flow applications is a common example of a large-scale inverse problem that involves the recovery of unknown interfaces. In the case of groundwater flow, for example, the inverse prob-

lem concerns the recovery of the interface between regions with different hydraulic conductivity given measurements of hydraulic head. Geometric inverse problems of this type have recently received a lot of attention by the research community Xie et al. (2011), Ping and Zhang (2014), Lorentzen et al. (2012), Lorentzen et al. (2012). Indeed, it has been recognized that the geometry determined by the aforementioned interfaces constitutes one of the main sources of uncertainty that must be quantified and reduced by means of Bayesian inversion.

In the context of groundwater flow, the identification of interfaces between regions associated with different types of geological properties can be posed as the recovery of a piecewise constant conductivity field parameterized with a level set function. A fully Bayesian level set framework for the solution of the aforementioned type of inverse problems has been recently developed in Iglesias et al. (2016). The MCMC method applied in Iglesias et al. (2016) performs well when the prior of the level set function properly encodes the intrinsic length scales of the unknown interfaces. Clearly, in practical applications such length scales are most likely unknown and their incorrect specification may result in inaccurate and uncertain estimates of the unknown interfaces. The purpose of this section is to show that the proposed hierarchical Bayesian framework enables us to determine an optimal length scale in the prior of the level set function which, in turn, captures more accurately the intrinsic length scale of the unknown interface.

4.3.1 The forward model

We are interested in the identification of a piecewise constant hydraulic conductivity, denoted by κ , of a two-dimensional confined aquifer whose physical domain is $D = [0, 6] \times [0, 6]$. We assume single-phase steady-state Darcy flow. The piezometric head, denoted by $h(x)$ ($x \in D$), which describes the flow within the aquifer can be modeled by the solution of Bear (1972)

$$-\nabla \cdot \kappa \nabla h = f \quad \text{in } D, \tag{12}$$

where f represents sources/sinks and where boundary conditions need to be specified. For the present work we consider the setup from the Benchmark used in Carrera and Neuman (1986), Hanke (1997), Iglesias and Dawson (2007), Iglesias et al. (2013), Iglesias (2016), and Iglesias et al. (2016). In concrete, we assume that f is a recharge term of the form

$$f(x_1, x_2) = \begin{cases} 0 & \text{if } 0 < x_2 \leq 4, \\ 137 & \text{if } 4 < x_2 < 5, \\ 274 & \text{if } 5 \leq x_2 < 6. \end{cases} \tag{13}$$

and we consider the following boundary conditions

$$\begin{aligned} h(x_1, 0) &= 100, & \frac{\partial h}{\partial x_1}(6, x_2) &= 0, \\ -\kappa \frac{\partial h}{\partial x_1}(0, x_2) &= 500, & \frac{\partial h}{\partial x_2}(x_1, 6) &= 0. \end{aligned} \tag{14}$$

We consider the inverse problem of recovering κ from observations $\{\ell_j(h)\}_{j=1}^{64}$ of h given by (12)–(14). We assume we have smoothed point observations given by

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

where $\varepsilon > 0$ and $\{q_j\}_{j=1}^{64} \subseteq D$ is a grid of 64 observation points equally distributed on D . Let $Z = L^p(D)$ for some $1 \leq p < \infty$ and $Y = \mathbb{R}^{64}$. Given $\kappa \in Z$, let h be given by (12)–(14). Then the forward map $S : Z \rightarrow Y$ is given by

$$\kappa \mapsto (\ell_1(h), \dots, \ell_{64}(h)).$$

We assume that each κ_i in the definition of the level set map F is strictly positive. The image of F is contained in the set of bounded fields on D bounded below by $n_i \kappa_i > 0$. In Iglesias et al. (2016) the map S is shown to be continuous and uniformly bounded on such fields, with respect to $\|\cdot\|_{L^p(D)}$ for some p , and so Assumptions 1 hold. As a consequence Theorem 2 applies directly.

4.3.2 Simulations and results

In the previous example we illustrate, with a simple model, the capabilities of the proposed framework to recover a specified true length scale and a true level set function that defines a true discontinuous field from which synthetic data are generated. However, we must reiterate that, in practice, we wish to recover the true discontinuous field; the level set function is merely an artifact that we use for the parameterization of such a field. In practical applications the aim of the proposed hierarchical Bayesian level set framework is to infer a length scale alongside with a level set function which, by means of expression (7), produces a discontinuous field that captures the desired piecewise constant field as accurately as possible and, in particular, the intrinsic length-scale separation of the interfaces determined by the discontinuities of the true geometric field. Therefore, in order to test our methodology in the applied setting of groundwater flow, rather than a true level set function, in this subsection we consider the true hydraulic conductivity κ^\dagger whose logarithm is displayed in Fig. 9a. This κ^\dagger is defined such that it takes the constant values $e^{1.5}$, e^4 and $e^{6.5}$. This is channelized conductivity typical of fluvial environments and often used as Benchmarks for subsurface flow inversion (Ping and Zhang 2014; Lorentzen et al.

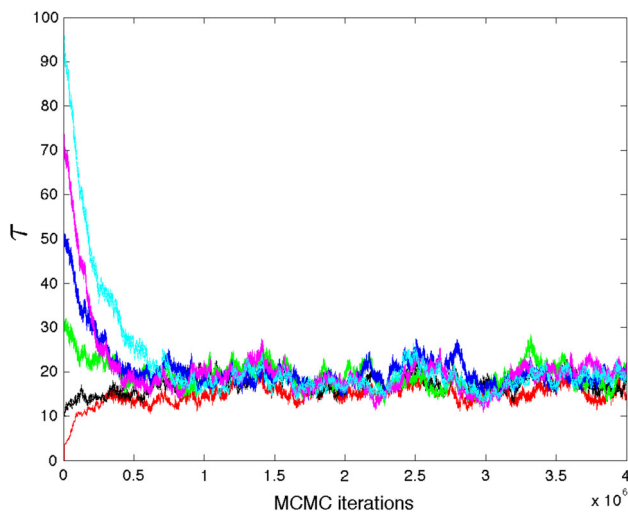


Fig. 8 (Groundwater flow model) trace plots of τ obtained from six hierarchical MCMC chains

2012; Xie et al. 2011; Iglesias et al. 2016). Note that the values that the conductivity can take on the three different regions differ by at least one order of magnitude, due to the logarithmic transformation. While there is indeed an intrinsic length scale in the channelized structure, this true conductivity field does not come from a specified level set prior.

Synthetic data are generated by means of

$$y = (\ell_1(h^\dagger), \dots, \ell_{64}(h^\dagger)) + \eta, \quad \eta \sim N(0, \Gamma) \text{ i.i.d.},$$

where h^\dagger is the solution to (12)–(14) for $\kappa = \kappa^\dagger$. Equations (12)–(14) have been solved with cell-centered finite differences Arbogast et al. (1997). In order to avoid inverse crimes, synthetic data are generated on a grid finer (160×160 cells) than the one used for the inversion (80×80 cells). The discretization is performed via the DFT, and we retain all modes. In addition, Γ is a diagonal matrix given by $\Gamma_{i,i} = 0.0175\ell_i(h^\dagger)$. In other words, we add noise that corresponds to 1.75% of the size of the noise-free observations. On the prior for the level set function u , we take Neumann boundary conditions and fix the smoothness parameter $\alpha = 5$.

We consider a Gaussian prior $N(35, 10^2)$ for τ , and use a Gaussian random walk proposal distribution for this parameter. We then apply the hierarchical MCMC method from subsection 3.3 initialized with the following six different choices of $\tau = 1, 10, 30, 50, 70, 90$ and a sample of the prior (with that given τ) of the level set function u . We thus produce six MCMC chains of length 4×10^6 and discard the first 10^6 as burn-in for the computation of quantities of interest. The trace plots of τ are displayed in Fig. 8 from which we clearly observe that all chains, regardless of their initial point, seem to stabilize and produce samples around $\tau = 18$. In the top row of Fig. 9b we display

the logarithm of some representatives samples of $F(u, \tau)$ under the hierarchical posterior. The middle row of Fig. 9b shows the logarithm of $F(\mathbb{E}(u), \mathbb{E}(\tau))$, i.e., the pushforward of the posterior means obtained using the hierarchical method. The bottom row of Fig. 9b displays the logarithm of the approximations of $\mathbb{E}(F(u, \tau))$. That is, the expected value of the pushforward samples under the posterior. The aforementioned results corresponds to five MCMC chains with τ initialized $\tau = 10, 30, 50, 70, 90$ (the results for $\tau = 1$ have been omitted). Similarly, Fig. 10 (top) shows the approximations of the variance of the pushforward samples of the posterior, i.e., $\text{Var}(F(u, \tau))$. Clearly, both $\mathbb{E}(F(u, \tau))$ and $F(\mathbb{E}(u), \mathbb{E}(\tau))$ result in fields that provide a reasonable approximation of the true geometric field. Note that, as expected, the largest uncertainty in the distribution of the pushforward samples is around the interface between the regions with different conductivity. In Fig. 11a we show some representative samples of u (top) and approximations to $\mathbb{E}(u)$ (bottom). In these plots, as before, we rescale the level set function by $\tau^{\alpha-d/2} = \tau^4$ so that they are all of approximately the same amplitude. In Fig. 12 we display the empirical densities of τ and the first five KL modes of u . A key observation is that, although the true hydraulic conductivity is not generated by thresholding a Gaussian random field, and hence there is no “true” length scale, the posterior nonetheless settles on a narrow range of values of τ which are consistent with the data.

From the aforementioned results we can also clearly see that the hierarchical MCMC algorithm produces similar outcomes regardless of the initialization of the inverse of the length scale τ , reflecting ergodicity of the Markov chain. The results from $\tau = 1$ are not shown but they are very similar to the ones from other chains. As with the results from the previous subsection, the similarity in outcomes between all six chains is not surprising as these are aimed at sampling from the same posterior distribution; but the fact that this posterior distribution on τ concentrates near to a single value is of particular interest because it shows that the true geometric field has an intrinsic length scale, even though it was not constructed via the map $F(u, \tau)$. Furthermore, this similarity of outcomes between chains showcases the main advantage of the proposed framework with respect to the non-hierarchical one. Indeed, as stated earlier, the proposed method has the ability to recover a distribution for the intrinsic length scale which gives rise to reasonably accurate estimates (i.e., $F(\mathbb{E}(u), \mathbb{E}(\tau))$, and $\mathbb{E}(F(u, \tau))$) of the true geometric field. We now present the numerical results from applying a non-hierarchical MCMC algorithm in which the inverse of length scale τ is fixed. We consider again six MCMC chains as before with the (now fixed) values of $\tau = 1, 10, 30, 50, 70, 90$ that we used to initialize the hierarchical chains used before. Analogous results to the ones presented for the hierarchical

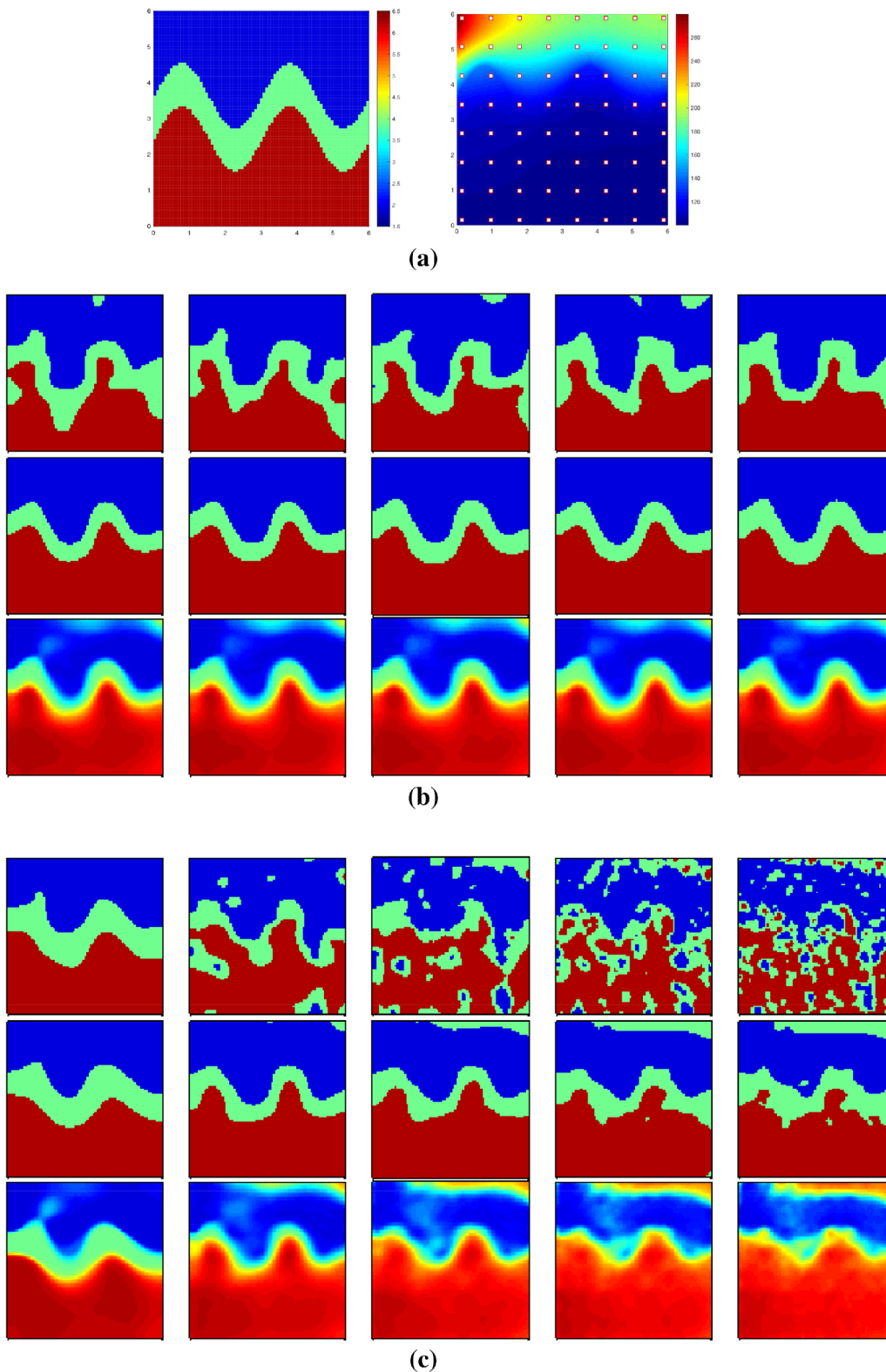


Fig. 9 Simulations for the groundwater flow model. **a** (Left) logarithm of the true hydraulic conductivity field used to generate the data y . (Right) true pressure field and the grid of observation points. **b** (Top) logarithm of representative samples of $F(u, \tau)$ under the hierarchical posterior. (Middle) logarithm of the approximations of $F(\mathbb{E}(u), \mathbb{E}(\tau))$.

(Bottom) logarithm of the approximations of $\mathbb{E}(F(u, \tau))$. From left-to-right, τ is initialized at $\tau = 10, 30, 50, 70, 90$. **c** As in **b**, using the non-hierarchical method. From left-to-right, τ is fixed at $\tau = 10, 30, 50, 70, 90$

Fig. 10 (Groundwater flow model) approximations of $\text{Var}(F(u, \tau))$ using the hierarchical (*top*) and the non-hierarchical (*bottom*) MCMC

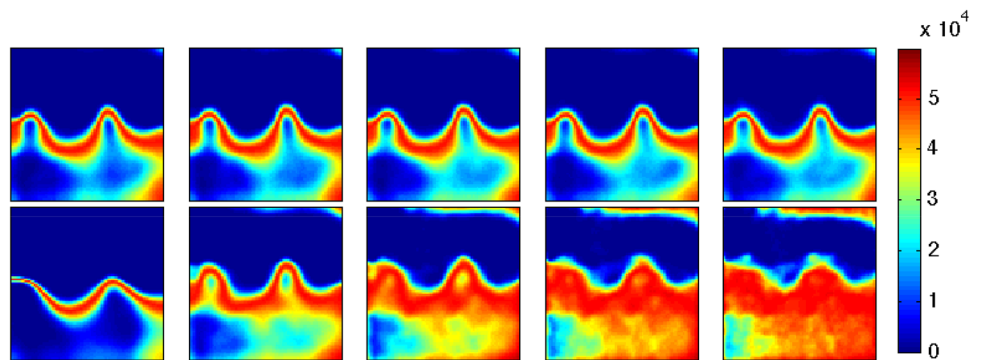
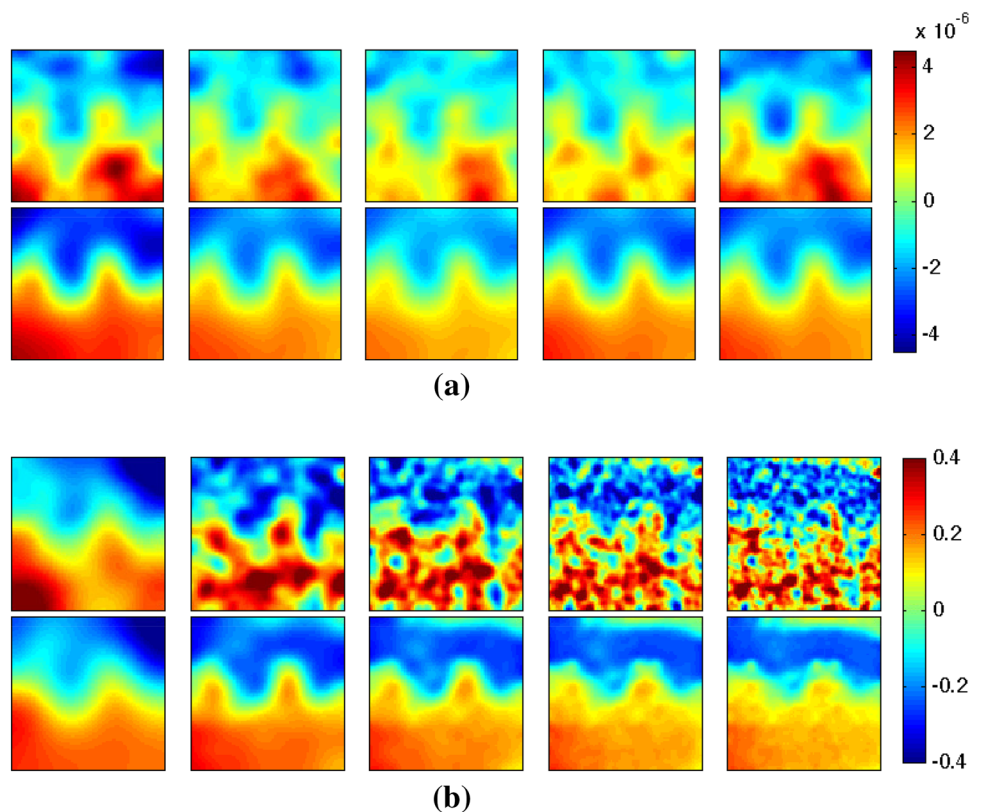


Fig. 11 (Groundwater flow model) representative samples and sample means of the level set function. The rescaling τ^4 means that the above quantities have the same approximate amplitude. True inverse length scale is $\tau = 15$. **a** (*Top*) representative samples of the rescaled level set function $\tau^4 \cdot u$ and (*bottom*) approximations of $\mathbb{E}(\tau^4 \cdot u)$ using the hierarchical method. From *left-to-right*, τ is initialized at $\tau = 10, 30, 50, 70, 90$. **b** As in **a**, using the non-hierarchical method. From *left-to-right*, τ is fixed at $\tau = 10, 30, 50, 70, 90$



method can be found in the bottom panels of Fig. 9 as well as the bottom of Figs. 10 and 11. Clearly, the lack of properly prescribing the intrinsic length scale in the non-hierarchical method results in inaccurate estimates of the true geometric field. We clearly observe that for $\tau \geq 30$ the estimates of the truth given by $F(\mathbb{E}(u), \mathbb{E}(\tau))$ and $\mathbb{E}(F(u, \tau))$ are substantially inaccurate and the uncertainty measured by $\text{Var}(F(u, \tau))$ is large. The non-hierarchical MCMC for $\tau = 1$ did not converge; the results are not shown. The non-hierarchical MCMC only provides reasonable estimates for $\tau = 10$ and $\tau = 30$. However, we can visually appreciate that these results are still suboptimal

when compared to the results from the hierarchical framework.

4.4 Electrical impedance tomography

Finally we consider the electrical impedance tomography (EIT) problem. This problem has previously been approached with a non-hierarchical Bayesian level set method Dunlop and Stuart (2015). In this subsection we show that the hierarchical approach outperforms the non-hierarchical approach in the case where the true conductivity is a binary field, given the same number of forward model evaluations.

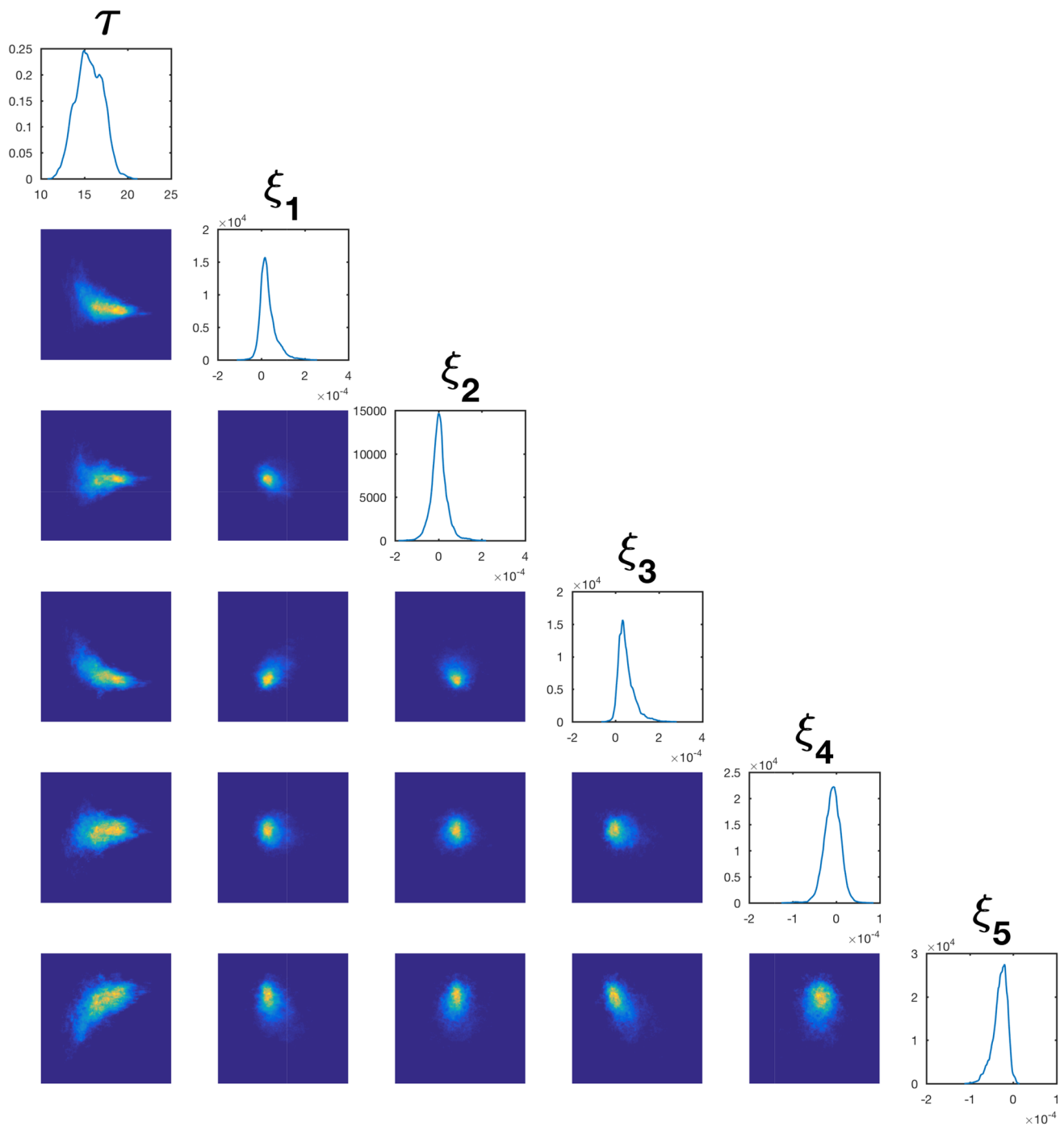


Fig. 12 (Groundwater flow model) (diagonal) empirical densities of τ and the first five KL modes of u . (off-diagonal) empirical joint densities

4.4.1 The forward model

EIT is an imaging technique which attempts to infer the internal conductivity of a body from boundary voltage measurements. Typical applications include medical imaging, as well as subsurface imaging where it is known as electrical

resistivity tomography (ERT). We utilize the complete electrode model (CEM), proposed in Somersalo et al. (1992). This is a physically accurate model which has been shown to agree with experimental data up to measurement precision. The strong form of the PDE governing the model is given by

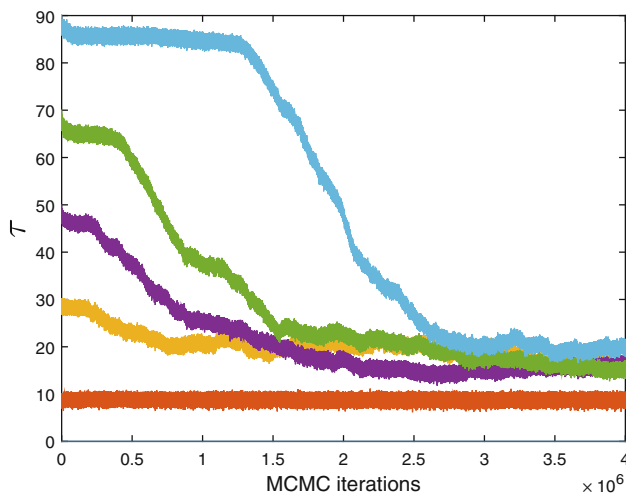


Fig. 13 (EIT model) the trace of τ along the MCMC chain, when initialized at the 5 different values $\tau = 10, 30, 50, 70, 90$

$$\begin{cases} -\nabla \cdot (\kappa(x) \nabla v(x)) = 0 & x \in D \\ \int_{e_l} \kappa \frac{\partial v}{\partial n} dS = I_l & l = 1, \dots, L \\ \kappa(x) \frac{\partial v}{\partial n}(x) = 0 & x \in \partial D \setminus \bigcup_{l=1}^L e_l \\ v(x) + z_l \kappa(x) \frac{\partial v}{\partial n}(x) = V_l & x \in e_l, l = 1, \dots, L. \end{cases}$$

Here $D \subseteq \mathbb{R}^2$ is the domain and $\{e_l\}_{l=1}^L \subseteq \partial D$ are electrodes on the boundary upon which currents $\{I_l\}_{l=1}^L$ are injected and voltages $\{V_l\}_{l=1}^L$ are read. The numbers $\{z_l\}_{l=1}^L$ represent the contact impedances of the electrodes. The field κ represents the conductivity of the body and v represents the potential within the body⁵. It should be noted that the solution of this PDE comprises both a potential $v \in H^1(D)$ and a vector $\{V_l\}_{l=1}^L$ of boundary voltage measurements.

The inverse problem we consider is the recovery of κ from a sequence of boundary voltage measurements. A number of (linearly independent) current stimulation patterns $\{I_l\}_{l=1}^L$ may be performed to provide more information; we assume that we perform the maximum $M = L - 1$ measurements. Let $Z = L^p(D)$ for some $1 \leq p < \infty$ and $Y = \mathbb{R}^J$ where $J = LM$. We can concatenate the boundary voltage measurements arising from different stimulation patterns to yield a map $S : Z \rightarrow Y$,

$$\kappa \mapsto (V^{(1)}, V^{(2)}, \dots, V^{(M)}),$$

$$\text{where } V^{(m)} = \{V_l^{(m)}\}_{l=1}^L \in \mathbb{R}^L, m = 1, \dots, M.$$

⁵ In the EIT literature the conductivity field is often denoted σ ; however, we have already used this in denoting the marginal variance of random fields.

For the experiments we work on a circular domain $D = \{x \in \mathbb{R}^2 \mid |x| < 1\}$. 16 electrodes are spaced equally around the boundary providing 50% coverage. All contact impedances are taken to be $z_l = 0.01$. Adjacent electrodes are stimulated with a current of 0.1, so that the matrix of stimulation patterns $I = \{I^{(j)}\}_{j=1}^{15} \in \mathbb{R}^{16 \times 15}$ is given by

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

We define our forward map $\mathcal{G} : X \times \mathbb{R}^+ \rightarrow \mathbb{R}^J$ by $\mathcal{G} = S \circ F$. As in the groundwater flow example, assume that each κ_i in the definition of the level set map is strictly positive. We do not have a continuity result for the map S on L^p for any $1 \leq p < \infty$. However, the almost-sure continuity of the map \mathcal{G} can be seen via a modification of the proof of Proposition 3.5 in Dunlop and Stuart (2015) to include the parameter τ ; this modification is almost identical to the proof of Proposition 1 given in the appendix. The uniform boundedness of \mathcal{G} follows from a result in Dunlop and Stuart (2015) similarly. Hence as was the case with the identity map example, the conclusions of Proposition 1 follow, and we can deduce the conclusions of Theorem 2.

4.4.2 Simulations and results

We fix a true conductivity κ^\dagger , shown in Fig. 14. As with the groundwater flow experiments, this is constructed explicitly and does not have a true value of τ associated with it. We generate data y as

$$y = S(\kappa^\dagger) + \eta, \quad \eta \sim N(0, \Gamma),$$

where we take the noise covariance $\Gamma = 0.0002^2 \cdot I$ to be white. The mean relative error on the generated data is approximately 12%. The data are generated using a mesh of 43264 elements and simulations are performed using a mesh of 10816 elements, in order to avoid an inverse crime. Forward solves are performed using the EIDORS software Adler and Lionheart (2006). All level set field samples are defined on the square $[-1, 1]^2$ and restricted to the domain D . This has the advantage of allowing for efficient sampling via the Fast Fourier Transform, though it has the drawback of introducing possibly non-trivial boundary effects on the domain; no such effects are observed in our problem, however. The discretization on the square is performed via the DFT on a grid of $2^7 \times 2^7$ points, and we retain all modes.

The level set map F is defined such that there are 2 phases, taking the constant values 1 and 10. We take the

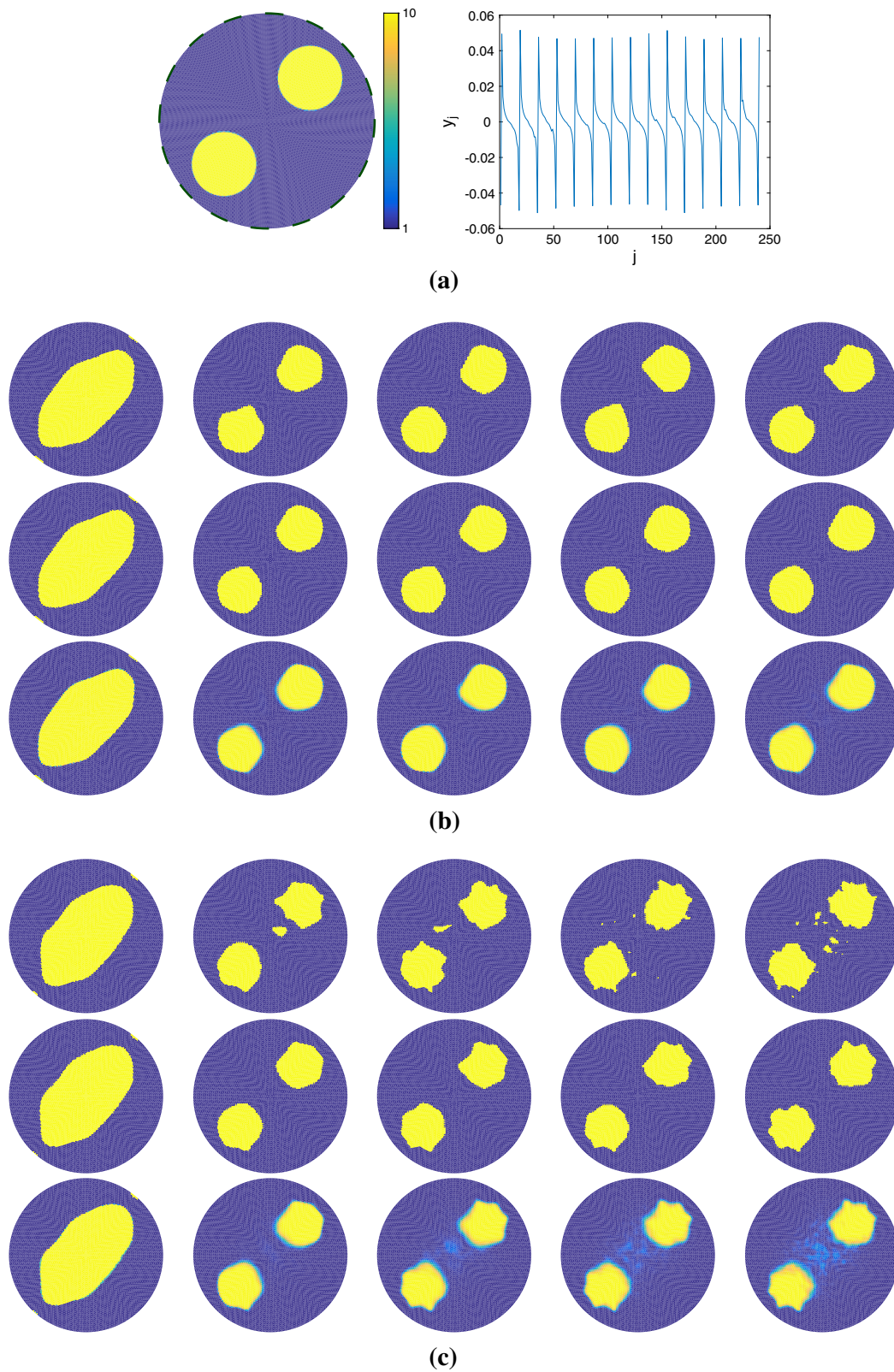


Fig. 14 Simulations for the EIT model. **a** (Left) true conductivity field used to generate the data y . (Right) the entries y_i of the data vector y , plotted against i . **b** (Top) Representative samples of $F(u, \tau)$ under the hierarchical posterior. (Middle) approximations of $F(\mathbb{E}(u), \mathbb{E}(\tau))$.

(Bottom) approximations of $\mathbb{E}(F(u, \tau))$. From left-to-right, τ is initialized at $\tau = 10, 30, 50, 70, 90$. **c** As in **b**, using the non-hierarchical method. From left-to-right, τ is fixed at $\tau = 10, 30, 50, 70, 90$

Fig. 15 (EIT model) approximations of $\text{Var}(F(u, \tau))$ using the hierarchical (*top*) and fixed (*bottom*) priors, with τ initialized or fixed at $\tau = 10, 30, 50, 70, 90$, from *left-to-right*

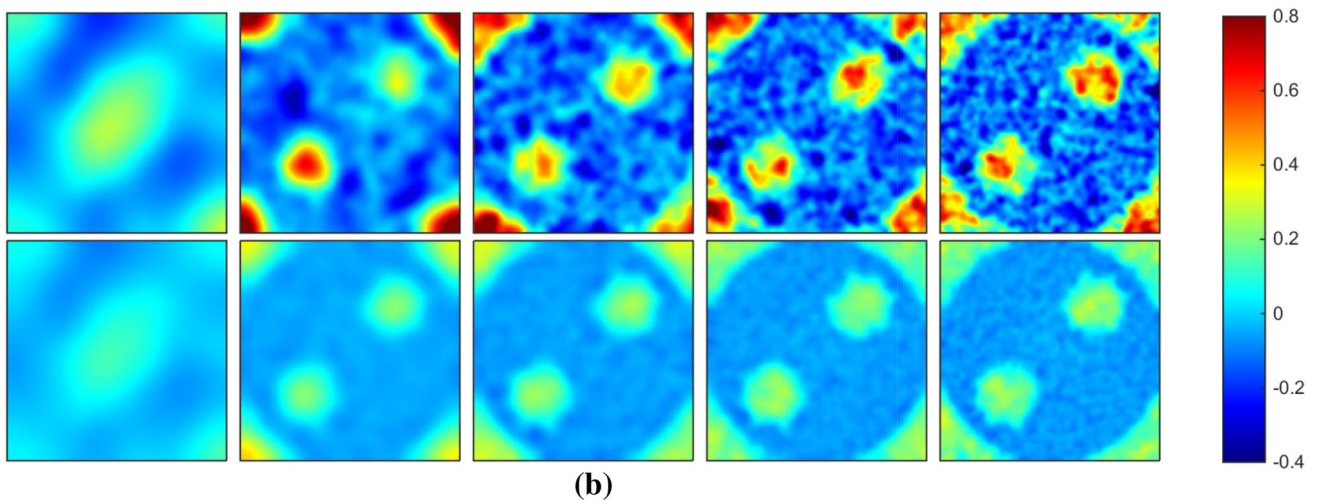
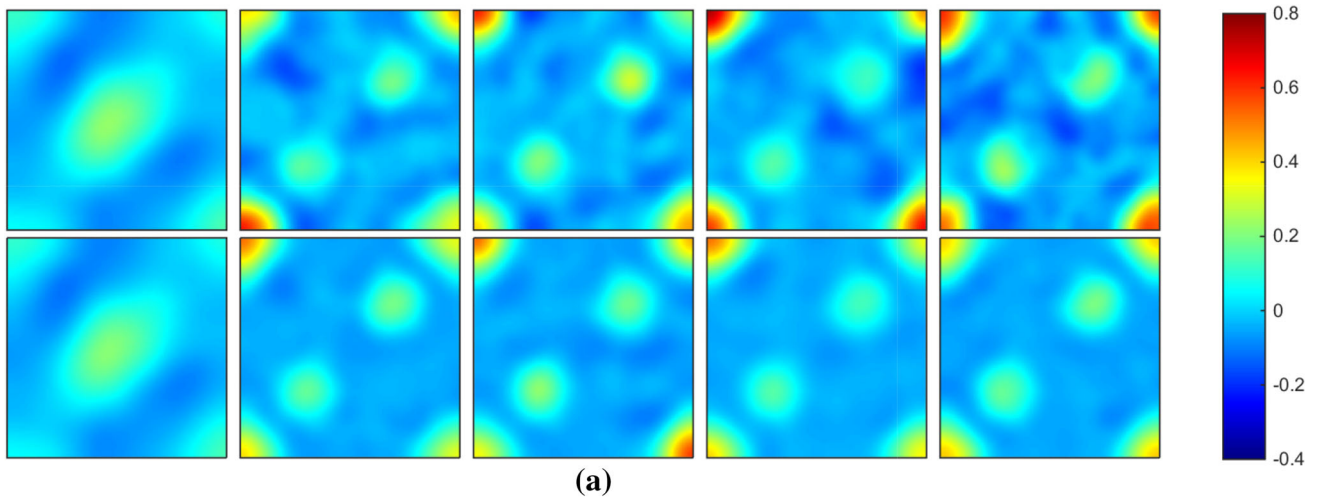
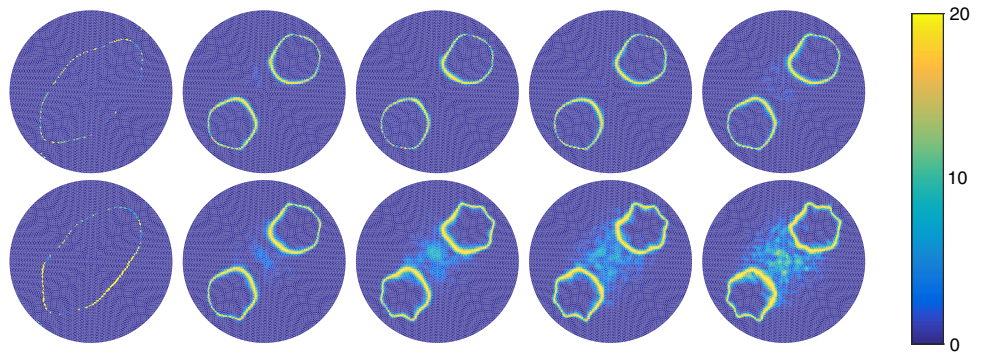


Fig. 16 (EIT model) representative samples and sample means of the level set function. The rescaling τ^4 means that the above quantities have the same approximate amplitude. True inverse length scale is $\tau = 15$. **a** (*Top*) representative samples of the rescaled level set function $\tau^4 \cdot u$ and

(*bottom*) approximations of $\mathbb{E}(\tau^4 \cdot u)$ using the hierarchical method. From *left-to-right*, τ is initialized at $\tau = 10, 30, 50, 70, 90$. **b** As in **a**, using the non-hierarchical method. From *left-to-right*, τ is fixed at $\tau = 10, 30, 50, 70, 90$

prior level set field mean to be zero, so that in this case F (and hence Φ) becomes independent of τ . Thus a forward model evaluation is not required for the Gibbs update of τ ,

and each sample of (u, τ) using the hierarchical method costs virtually the same as one of u using the non-hierarchical method.

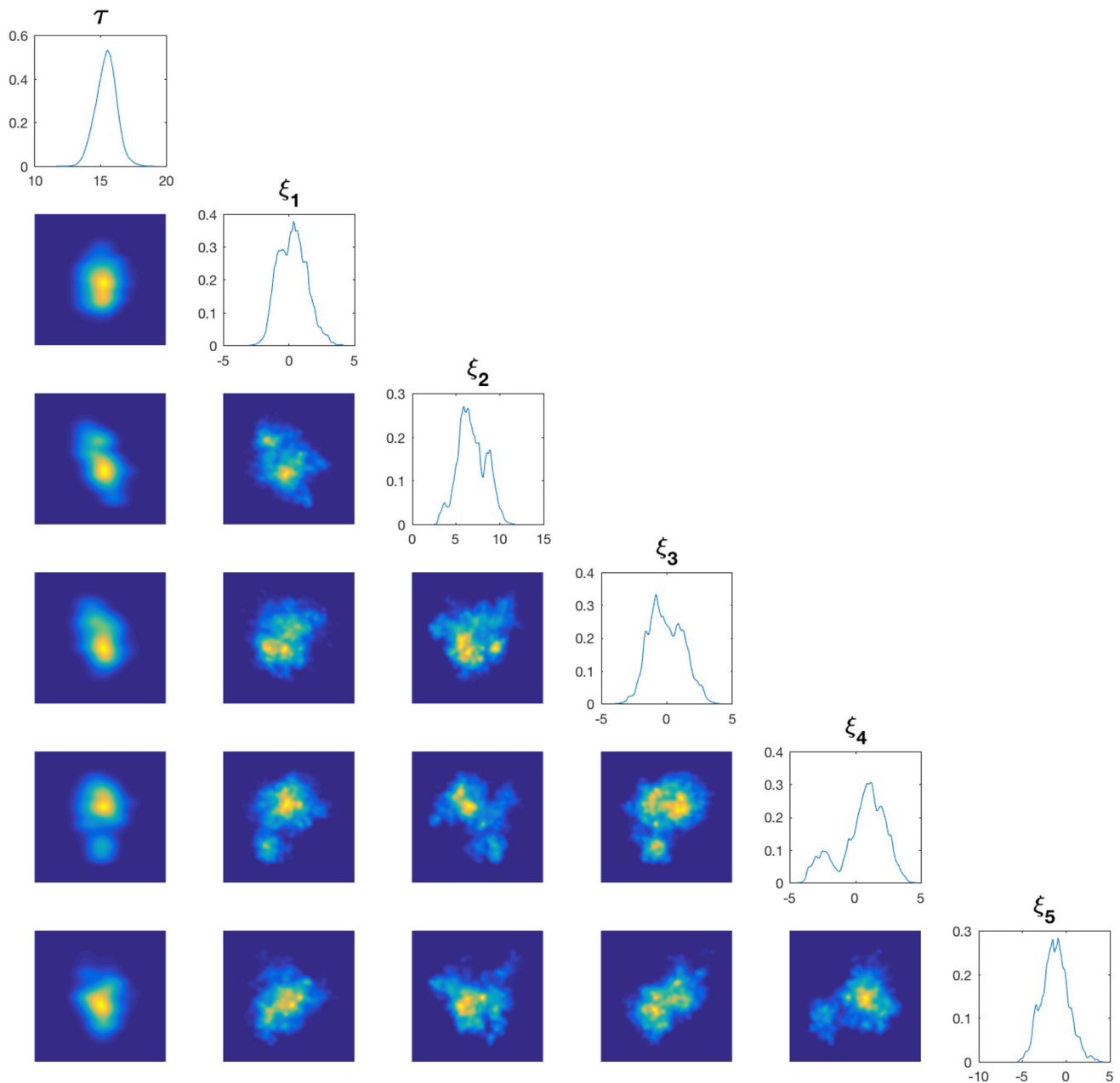


Fig. 17 (EIT model) (diagonal) empirical densities of τ and the first five KL modes of u . (off-diagonal) empirical joint densities

Similarly to the previous experiments, we initialize the hierarchical sampling from $\tau = 10, 30, 50, 70, 90$ to check for robustness of the method. We use a sharper prior on τ than was used previously. We again use a Gaussian random walk proposal distribution for τ . We fix the smoothness parameter $\alpha = 5$ in the prior for u , and again use Neumann boundary conditions. We again wish to compare how the hierarchical method compares with the non-hierarchical method. We therefore also look at the 5 different posterior distributions that arise when using each of 5 fixed prior inverse length scales $\tau = 10, 30, 50, 70, 90$, which gives another 5 MCMC

chains. For both the methods we produce 4×10^6 samples for each chain, and discard the first 2×10^6 samples as burn-in when calculating quantities of interest.

The traces of the values of τ along the hierarchical chains are shown in Fig. 13. With the exception of the chain initialized at $\tau = 10$, the chains converge to the sample approximate value of τ . Unlike in previous experiments, the traces have a relatively flat period before the approximate linear convergence to the common length scale. Initializing $\tau = 90$ requires an additional 10^6 samples to converge, over the other converging chains.

Figure 14 shows the push forwards of the sample means from different chains under the level set map, along with approximations of $\mathbb{E}(F(u, \tau))$ and typical samples of $F(u, \tau)$ coming from the different posteriors. In both the hierarchical and non-hierarchical methods, the chains initialized/fixed at $\tau = 10$ fail to recover the true conductivity, similarly to what was observed with the identity map experiments when initializing at $\tau = 5$. The other chains for the hierarchical method produce very similar results to one another, while the effect of fixing the length scale to be too short is apparent in the figures for the non-hierarchical method.

In Fig. 15 we see approximations to $\text{Var}(F(u, \tau))$ under the different posteriors. In both cases, variance is highest around the boundaries of the two inclusions. The difference between the hierarchical and non-hierarchical methods is more apparent here, with higher variance between the two inclusions when the length scale is fixed to be too short.

Again, we look at the level set function u itself in Fig. 16. In these plots, as before, we rescale the level set function by $\tau^{\alpha-d/2} = \tau^4$ so that they are all of approximately the same amplitude. As in the previous experiments, there is noticeable contrast between the means for the hierarchical and non-hierarchical methods, and yet more contrast between the typical samples.

Finally, in Fig. 17, we show the posterior densities on the inverse length scale and the first five KL modes, as well as correlations between them. As with the groundwater flow example, although there is no “true” inverse length scale, the data are sufficiently informative to define a small range of values for this parameter under the posterior.

5 Conclusions

The level set method is an attractive approach to inverse problems for the detection of interfaces. Furthermore, the Bayesian approach is particularly desirable when there is a need to quantify uncertainty. In this paper we have shown that Bayesian level set inversion is considerably enhanced by a hierarchical approach in which the length scale of the underlying level set function is inferred from the data. We have demonstrated this by means of three examples of interest arising in, respectively, the information, physical and medical sciences; however, many potential applications remain to be explored and this provides an interesting avenue for future work.

We also developed the theoretical underpinnings for our hierarchical method. Our work is based on a Metropolis-within-Gibbs approach which alternates between updating the level set function and the length scale. The Metropolis method we use for the level set field update does not use derivatives of the log-likelihood, and could be improved by doing so, using the infinite-dimensional variants on MALA

and HMC (which use first derivative information, see the citations in Cotter et al. (2013)) or the manifold MALA and HMC methods, which use higher order derivatives Girolami and Calderhead (2011). Another interesting direction for future work is the design of methods with more informed proposals which exploit correlations in the level set function and its length scale. And finally it would be interesting to consider pseudo-marginal methods to sample the hierarchical parameter alone, as in Filippone and Girolami (2014).

Assuming independence under the prior, it would require little further work to treat the thresholding levels $\{c_i\}$ and the values of the thresholded function $\{\kappa_i\}$ as part of the inference as well; we omitted this here for the sake of clarity. Such a model may be more realistic, and numerical studies of such models may prove interesting. Another extension of interest may be to place a hyperprior upon the regularity parameter also, which may be useful for improving rates of convergence van der Vaart and van Zanten (2009). This is a more challenging task, again related to singularity of measures. The paper Agapiou et al. (2014) discusses ways in which this may be done; however, it is still an open question in terms of theory and optimal algorithms. Additionally, it may be of interest to overcome the restriction of the ordering of phases $\{\kappa_i\}$ by means of a vector level set method Tai and Chan (2004).

Finally we mention that the use of a single length scale within an isotropic prior is a simple example of more sophisticated hierarchical approaches which attempt to learn non-stationary and non-isotropic Calvetti and Somersalo (2007), Calvetti and Somersalo (2008) features of the level set function from the data. This provides an interesting opportunity for future work and for ideas from machine learning to play a role in the solution of inverse problems for interfaces.

Acknowledgments AMS is grateful to DARPA, EPSRC, and ONR for the financial support. MMD was supported by the EPSRC-funded MASDOC graduate training program. The authors are grateful to Dan Simpson for helpful discussions. The authors are also grateful for discussions with Omiros Papaspiliopoulos about links with probit. The authors would also like to thank the two anonymous referees for their comments that have helped improve the quality of the paper. This research utilized Queen Mary’s MidPlus computational facilities, supported by QMUL Research-IT and funded by EPSRC grant EP/K000128/1.

Appendix

Proof of theorems

Proof (Theorem 1)

- (i) Note that it suffices to show that $\mu_0^\tau \sim \mu_0^0$ for all $\tau > 0$. (Here \sim denotes “equivalent as measures”). It is known

that the eigenvalues of $-\Delta$ on \mathbb{T}^d grow like $j^{2/d}$, and hence the eigenvalues $\lambda_j(\tau)$ of $\mathcal{C}_{\alpha,\tau}$ decay like

$$\lambda_j(\tau) \asymp (\tau^2 + j^{2/d})^{-\alpha}, \quad j \geq 1.$$

Using Proposition 3 below, we see that $\mu_0^\tau \sim \mu_0^0$ if

$$\sum_{j=1}^{\infty} \left(\frac{\lambda_j(\tau)}{\lambda_j(0)} - 1 \right)^2 < \infty.$$

Now we have

$$\begin{aligned} \left| \frac{\lambda_j(\tau)}{\lambda_j(0)} - 1 \right| &\asymp \left| \left(1 + \frac{\tau^2}{j^{2/d}} \right)^{-\alpha} - 1 \right| \\ &\leq \left| \exp\left(\frac{\alpha\tau^2}{j^{2/d}} \right) - 1 \right| \\ &\leq C \frac{\alpha\tau^2}{j^{2/d}}. \end{aligned}$$

Here we have used that $(1+x)^{-\alpha} - 1 \leq \exp(\alpha x) - 1$ for all $x \geq 0$ to move from the first to the second line, and that $\exp(x) - 1 \leq Cx$ for all $x \in [0, x_0]$ to move from the second to third line. Now note that when $d \leq 3$, $j^{-4/d}$ is summable, and so it follows that $\mu_0^\tau \sim \mu_0^0$.

- (ii) The case $\tau = 0$ is Theorem 2.18 in Dashti and Stuart (2016); the general result follows from the equivalence above.
- (iii) Let $v \sim N(0, \mathcal{D}_{\sigma,v,\ell})$ where $\mathcal{D}_{\sigma,v,\ell}$ is as given by (2). Then we have

$$\begin{aligned} \mathcal{D}_{\sigma,v,\ell} &= \beta \ell^d (I - \ell^2 \Delta)^{-v-d/2} \\ &= \beta \ell^d \ell^{-2v-d} (\ell^{-2} I - \Delta)^{-v-d/2} \\ &= \beta \tau^{2\alpha-d} (\tau^2 I - \Delta)^{-\alpha} \\ &= \beta \tau^{2\alpha-d} \mathcal{C}_{\alpha,\tau}. \end{aligned}$$

Hence, letting $u \sim N(0, \mathcal{C}_{\alpha,\tau})$, we see that

$$\begin{aligned} \mathbb{E}\|u\|^2 &= \text{tr}(\mathcal{C}_{\alpha,\tau}) \\ &= \frac{1}{\beta} \tau^{d-2\alpha} \text{tr}(\mathcal{D}_{\sigma,v,\ell}) \\ &= \frac{1}{\beta} \tau^{d-2\alpha} \mathbb{E}\|v\|^2. \end{aligned}$$

□

Proof (Theorem 2) Proposition 1 which follows shows that μ_0 and Φ satisfy Assumptions 2.1 in Iglesias et al. (2016), with $U = X \times \mathbb{R}^+$. Theorem 2.2 in Iglesias et al. (2016) then tells us that the posterior exists and is Lipschitz with respect to the data. □

Proposition 1 Let μ_0 be given by (3) and $\Phi : X \times \mathbb{R}^+ \rightarrow \mathbb{R}$ be given by (10). Let Assumptions 1 hold. Then

- (i) for every $r > 0$ there is a $K = K(r)$ such that, for all $(u, \tau) \in X \times \mathbb{R}^+$ and all $y \in Y$ with $|y|_r < r$,

$$0 \leq \Phi(u, \tau; y) \leq K;$$

- (ii) for any fixed $y \in Y$, $\Phi(\cdot, \cdot; y) : X \times \mathbb{R}^+ \rightarrow \mathbb{R}$ is continuous μ_0 -almost surely on the complete probability space $(X \times \mathbb{R}^+, \mathcal{X} \otimes \mathcal{R}, \mu_0)$;
- (iii) for $y_1, y_2 \in Y$ with $\max\{|y_1|_r, |y_2|_r\} < r$, there exists a $C = C(r)$ such that for all $(u, \tau) \in X \times \mathbb{R}^+$,

$$|\Phi(u, \tau; y_1) - \Phi(u, \tau; y_2)| \leq C|y_1 - y_2|_r.$$

Proof (i) Recall the level set map F defined by (7) defined via the finite constant values κ_i taken on each subset D_i of \bar{D} . We may bound F uniformly:

$$|F(u, \tau)| \leq \max\{|\kappa_1|, \dots, |\kappa_n|\} =: F_{\max},$$

for all $(u, \tau) \in X \times \mathbb{R}^+$. Combining this with Assumption 1(ii), it follows that \mathcal{G} is uniformly bounded on $X \times \mathbb{R}^+$. The result then follows from the continuity of $y \mapsto \frac{1}{2}|y - \mathcal{G}(u, \tau)|_r^2$.

- (ii) Let $(u, \tau) \in X \times \mathbb{R}^+$ and let $D_i(u, \tau)$ be as defined by (6), and define $D_i^0(u, \tau)$ by

$$\begin{aligned} D_i^0(u, \tau) &= \bar{D}_i(u, \tau) \cap \bar{D}_{i+1}(u, \tau) \\ &= \{x \in D \mid u(x) = c_i(\tau)\}, \quad i = 1, \dots, n-1. \end{aligned}$$

We first show that \mathcal{G} is continuous at (u, τ) whenever $|D_i^0(u, \tau)| = 0$ for $i = 1, \dots, n-1$. Choose an approximating sequence $\{u_\varepsilon, \tau_\varepsilon\}_{\varepsilon>0}$ of (u, τ) such that $\|u_\varepsilon - u\|_\infty + |\tau_\varepsilon - \tau| < \varepsilon$ for all $\varepsilon > 0$. We will first show that $\|F(u_\varepsilon, \tau_\varepsilon) - F(u, \tau)\|_{L^p(D)} \rightarrow 0$ for any $p \in [1, \infty)$. As in Iglesias et al. (2016) Proposition 2.4, we can write

$$\begin{aligned} F(u_\varepsilon, \tau_\varepsilon) - F(u, \tau) &= \sum_{i=1}^n \sum_{j=1}^n (\kappa_i - \kappa_j) \mathbb{1}_{D_i(u_\varepsilon, \tau_\varepsilon) \cap D_j(u, \tau)} \\ &= \sum_{\substack{i,j=1 \\ i \neq j}}^n (\kappa_i - \kappa_j) \mathbb{1}_{D_i(u_\varepsilon, \tau_\varepsilon) \cap D_j(u, \tau)}. \end{aligned}$$

From the definition of $(u_\varepsilon, \tau_\varepsilon)$,

$$u(x) - \varepsilon < u_\varepsilon(x) < u(x) + \varepsilon, \quad \tau - \varepsilon < \tau_\varepsilon < \tau + \varepsilon$$

for all $x \in D$ and $\varepsilon > 0$. We claim that for $|i - j| > 1$ and ε sufficiently small, $D_i(u_\varepsilon, \tau_\varepsilon) \cap D_j(u, \tau) = \emptyset$.

First note that

$$D_i(u_\varepsilon, \tau_\varepsilon) = \{x \in D \mid \tau_\varepsilon^{d/2-\alpha} c_{i-1} \leq u_\varepsilon(x) < \tau_\varepsilon^{d/2-\alpha} c_i\}$$

$$= \{x \in D \mid c_{i-1} \leq \tau_\varepsilon^{\alpha-d/2} u_\varepsilon(x) < c_i\}.$$

$$\leq (2F_{\max})^p \sum_{\substack{i,j=1 \\ i \neq j}}^n |D_i(u_\varepsilon, \tau_\varepsilon) \cap D_j(u, \tau)|$$

$$\rightarrow 0.$$

Then we have that

$$D_i(u_\varepsilon, \tau_\varepsilon) \cap D_j(u, \tau)$$

$$= \{x \in D \mid c_{i-1} \leq \tau_\varepsilon^{\alpha-d/2} u_\varepsilon(x) < c_i,$$

$$c_{j-1} \leq \tau^{\alpha-d/2} u(x) < c_j\}.$$

Now, since u is bounded,

$$\tau^{\alpha-d/2} u(x) - \mathcal{O}(\varepsilon) < \tau_\varepsilon^{\alpha-d/2} u_\varepsilon(x)$$

$$< \tau^{\alpha-d/2} u(x) + \mathcal{O}(\varepsilon)$$

and so

$$D_i(u_\varepsilon, \tau_\varepsilon) \cap D_j(u, \tau) \subseteq$$

$$\{x \in D \mid c_{i-1} - \mathcal{O}(\varepsilon) \leq \tau^{\alpha-d/2} u(x) < c_i + \mathcal{O}(\varepsilon),$$

$$c_{j-1} \leq \tau^{\alpha-d/2} u(x) < c_j\}.$$

From the strict ordering of the $\{c_i\}_{i=1}^n$ we deduce that for $|i - j| > 1$ and small enough ε , the right-hand side is empty. We hence look at the cases $|i - j| = 1$. With the same reasoning as above, we see that

$$D_i(u_\varepsilon, \tau_\varepsilon) \cap D_{i+1}(u, \tau)$$

$$\subseteq \{x \in D \mid c_i - \mathcal{O}(\varepsilon) \leq \tau^{\alpha-d/2} u(x) < c_i + \mathcal{O}(\varepsilon)\}$$

$$\rightarrow \{x \in D \mid \tau^{\alpha-d/2} u(x) = c_i\}$$

$$= \{x \in D \mid u(x) = \tau^{d/2-\alpha} c_i\}$$

$$= D_i^0(u, \tau)$$

and also

$$D_i(u_\varepsilon, \tau_\varepsilon) \cap D_{i-1}(u, \tau)$$

$$\subseteq \{x \in D \mid c_{i-1} - \mathcal{O}(\varepsilon) < \tau^{\alpha-d/2} u(x) < c_{i-1}\}$$

$$\rightarrow \emptyset.$$

Assume that each $|D_i^0(u, \tau)| = 0$, then it follows that $|D_i(u_\varepsilon, \tau_\varepsilon) \cap D_j(u, \tau)| \rightarrow 0$ whenever $i \neq j$. Therefore we have that

$$\|F(u_\varepsilon, \tau_\varepsilon) - F(u, \tau)\|_{L^p(D)}^p$$

$$= \sum_{\substack{i,j=1 \\ i \neq j}}^n \int_{D_i(u_\varepsilon, \tau_\varepsilon) \cap D_j(u, \tau)} |\kappa_i - \kappa_j|^p dx$$

Thus F is continuous at (u, τ) . By Assumption 1(i) it follows that \mathcal{G} is continuous at (u, τ) .

We now claim that $|D_i^0(u, \tau)| = 0$ μ_0 -almost surely for each i . By Tonelli's theorem, we have that

$$\mathbb{E}|D_i^0(u, \tau)|$$

$$= \int_{X \times \mathbb{R}^+} |D_i^0(u, \tau)| \mu_0(du, d\tau)$$

$$= \int_{X \times \mathbb{R}^+} \left(\int_{\mathbb{R}} \mathbb{1}_{D_i^0(u, \tau)}(x) dx \right) \mu_0(du, d\tau)$$

$$= \int_{\mathbb{R}^d} \left(\int_{X \times \mathbb{R}^+} \mathbb{1}_{D_i^0(u, \tau)}(x) \mu_0(du, d\tau) \right) dx$$

$$= \int_{\mathbb{R}^d} \left(\int_0^\infty \left(\int_X \mathbb{1}_{D_i^0(u, \tau)}(x) \mu_0^\tau(du) \right) \pi_0(d\tau) \right) dx$$

$$= \int_{\mathbb{R}^d} \left(\int_0^\infty \mu_0^\tau(\{u \in X \mid u(x) = c_i(\tau)\}) \pi_0(d\tau) \right) dx.$$

For each $\tau \geq 0$ and $x \in D$, $u(x)$ is a real-valued Gaussian random variable under μ_0^τ . It follows that $\mu_0^\tau(\{u \in X \mid u(x) = c_i(\tau)\}) = 0$, and so $\mathbb{E}|D_i^0(u, \tau)| = 0$. Since $|D_i^0(u, \tau)| \geq 0$ we have that $|D_i^0(u, \tau)| = 0$ μ_0 -almost surely. The result now follows.

(iii) For fixed $(u, \tau) \in X \times \mathbb{R}^+$, the map $y \mapsto \frac{1}{2}|y - \mathcal{G}(u, \tau)|_F^2$ is smooth and hence locally Lipschitz. \square

Proof (Theorem 4) Recall that the eigenvalues of $\mathcal{C}_{\alpha, \tau}$ satisfy $\lambda_j(\tau) \asymp (\tau^2 + j^{2/d})^{-\alpha}$. Then we have that

$$\left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1 \right) \asymp (1 + \tau^2 j^{-2/d})^\alpha - 1 = \mathcal{O}(j^{-2/d}).$$

It follows that

$$\sum_{j=1}^\infty \left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1 \right)^p < \infty \text{ if and only if } d < 2p. \quad (15)$$

(i) We first prove the ‘if’ part of the statement. We have $u \sim N(0, \mathcal{C}_0)$, and so $\mathbb{E}\langle u, \varphi_j \rangle^2 = \lambda_j(0)$. Since the terms within the sum are non-negative, by Tonelli's theorem we can bring the expectation inside the sum to see that

$$\mathbb{E} \sum_{j=1}^\infty \left(\frac{1}{\lambda_j(\tau)} - \frac{1}{\lambda_j(0)} \right) \langle u, \varphi_j \rangle^2 = \sum_{j=1}^\infty \left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1 \right)$$

which is finite if and only if $d < 2$, i.e., $d = 1$. It follows that the sum is finite almost surely.

For the converse, suppose that $d \geq 2$ so that the series in (15) diverges when $p = 1$. Let $\{\xi_j\}_{j \geq 1}$ be a sequence of i.i.d. $N(0, 1)$ random variables so that $\langle u, \varphi_j \rangle^2$ has the same distribution as $\lambda_j(0)\xi^2$. Define the sequence $\{Z_n\}_{n \geq 1}$ by

$$\begin{aligned} Z_n &= \sum_{j=1}^n \left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1 \right) \xi_j^2 \\ &= \sum_{j=1}^n \left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1 \right) + \sum_{j=1}^n \left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1 \right) (\xi_j^2 - 1) \\ &=: X_n + Y_n. \end{aligned}$$

Then the result follows if Z_n diverges with positive probability. By assumption we have that X_n diverges. In order to show that Z_n diverges with positive probability it hence suffices to show that Y_n converges with positive probability. Define the sequence of random variables $\{W_j\}_{j \geq 1}$ by

$$W_j = \left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1 \right) (\xi_j^2 - 1).$$

It can be checked that

$$\mathbb{E}(W_j) = 0, \quad \text{Var}(W_j) = 2 \left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1 \right)^2.$$

The series of variances converges if and only if $d \leq 3$, using (15) with $p = 2$. We use Kolmogorov’s two series theorem, Theorem 3.11 in Srinivasa Varadhan (2001), to conclude that $Y_n = \sum_{j=1}^n W_j$ converges almost surely and the result follows.

(ii) Now we have

$$\begin{aligned} \log \left(\frac{\lambda_j(\tau)}{\lambda_j(0)} \right) &= -\log \left(1 - \left(1 - \frac{\lambda_j(0)}{\lambda_j(\tau)} \right) \right) \\ &= \left(1 - \frac{\lambda_j(0)}{\lambda_j(\tau)} \right) + \frac{1}{2} \left(1 - \frac{\lambda_j(0)}{\lambda_j(\tau)} \right)^2 \\ &\quad + \text{h.o.t.} \end{aligned}$$

Let $\{\xi_j\}_{j \geq 1}$ be a sequence of i.i.d. $N(0, 1)$ random variables, so that again we have that $\langle u, \varphi_j \rangle^2$ has the same distribution as $\lambda_j(0)\xi^2$. Then it is sufficient to show that the series

$$I = \sum_{j=1}^{\infty} \left[\left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1 \right) \xi_j^2 + \log \left(\frac{\lambda_j(\tau)}{\lambda_j(0)} \right) \right]$$

is finite almost surely. We use the above approximation for the logarithm to write

$$\begin{aligned} I &= \sum_{j=1}^{\infty} \left(\frac{\lambda_j(0)}{\lambda_j(\tau)} - 1 \right) (\xi_j^2 - 1) \\ &\quad + \sum_{j=1}^{\infty} \left[\frac{1}{2} \left(1 - \frac{\lambda_j(0)}{\lambda_j(\tau)} \right)^2 + \text{h.o.t.} \right]. \end{aligned}$$

The second sum converges if and only if $d < 4$, i.e., $d \leq 3$. The almost-sure convergence of the first term is shown in the proof of part (i). \square

Proposition 2 Let $D \subseteq \mathbb{R}^d$. Define the construction map $F : X \times \mathbb{R}^+ \rightarrow \mathbb{R}^D$ by (7). Given $x_0 \in D$ define $\mathcal{G} : X \times \mathbb{R}^+ \rightarrow \mathbb{R}$ by $\mathcal{G}(u, \tau) = F(u, \tau)|_{x_0}$. Then \mathcal{G} is continuous at any $(u, \tau) \in X \times \mathbb{R}^+$ with $u(x_0) \neq c_i(\tau)$ for each $i = 0, \dots, n$. In particular, \mathcal{G} is continuous μ_0 -almost surely when μ_0 is given by (3). Additionally, \mathcal{G} is uniformly bounded.

Proof The uniform boundedness is clear. For the continuity, let $(u, \tau) \in X \times \mathbb{R}^+$ with $u(x_0) \neq c_i(\tau)$ for each $i = 0, \dots, n$. Then there exists a unique j such that

$$c_{j-1}(\tau) < u(x_0) < c_j(\tau). \tag{16}$$

Given $\delta > 0$, let $(u_\delta, \tau_\delta) \in X \times \mathbb{R}^+$ be any pair such that

$$\|u_\delta - u\|_\infty + |\tau_\delta - \tau| < \delta.$$

Then it is sufficient to show that for all δ sufficiently small, $x_0 \in D_j(u_\delta, \tau_\delta)$, i.e., that

$$c_{j-1}(\tau_\delta) \leq u_\delta(x_0) < c_j(\tau_\delta).$$

From this it follows that $G(u_\delta, \tau_\delta) = G(u, \tau)$.

Since the inequalities in (16) are strict, we can find $\alpha > 0$ such that

$$c_{j-1} + \alpha < u(x_0) < c_j(\tau) - \alpha. \tag{17}$$

Now c_j is continuous at $\tau > 0$, and so there exists a $\gamma > 0$ such that for any $\lambda > 0$ with $|\lambda - \tau| < \gamma$ we have

$$c_j(\lambda) - \alpha/2 < c_j(\tau) < c_j(\lambda) + \alpha/2. \tag{18}$$

We have that $\|u_\delta - u\|_\infty < \delta$, and so in particular,

$$u(x_0) - \delta < u_\delta(x_0) < u(x_0) + \delta. \tag{19}$$

We can combine (17)–(19) to see that, for $\delta < \gamma$,

$$c_{j-1}(\tau_\delta) - \delta + \alpha/2 < u_\delta(x_0) < c_j(\tau_\delta) + \delta - \alpha/2$$

and so in particular, for $\delta < n\{\gamma, \alpha/2\}$,

$$c_{j-1}(\tau_\delta) < u_\delta(x_0) < c_j(\tau_\delta).$$

\square

Radon–Nikodym derivatives in Hilbert spaces

The following proposition gives an explicit formula for the density of one Gaussian with respect to another and is used in defining the acceptance probability for the length-scale updates in our algorithm. Although we only use the proposition in the case where H is a function space and the mean m is zero, we provide a proof in the more general case where m is an arbitrary element of separable Hilbert space H as this setting may be of independent interest.

Proposition 3 *Let $(H, \langle \cdot, \cdot \rangle, \|\cdot\|)$ be a separable Hilbert space, and let A, B be positive trace-class operators on H . Assume that A and B share a common complete set of orthonormal eigenvectors $\{\varphi_j\}_{j \geq 1}$, with the eigenvalues $\{\lambda_j\}_{j \geq 1}$, $\{\gamma_j\}_{j \geq 1}$ defined by*

$$A\varphi_j = \lambda_j\varphi_j, \quad B\varphi_j = \gamma_j\varphi_j,$$

for all $j \geq 1$. Assume further that the eigenvalues satisfy

$$\sum_{j=1}^{\infty} \left(\frac{\lambda_j}{\gamma_j} - 1 \right)^2 < \infty.$$

Let $m \in H$ and define the measures $\mu = N(m, A)$ and $\nu = N(m, B)$. Then μ and ν are equivalent, and their Radon–Nikodym derivative is given by

$$\frac{d\mu}{d\nu}(u) = \prod_{j=1}^{\infty} \frac{\gamma_j}{\lambda_j} \cdot \exp\left(\frac{1}{2} \sum_{j=1}^{\infty} \left(\frac{1}{\gamma_j} - \frac{1}{\lambda_j}\right) \langle u - m, \varphi_j \rangle^2\right).$$

Proof The assumption on summability of the eigenvalues means that the Feldman–Hájek theorem applies, and so we know that μ and ν are equivalent. We show that the Radon–Nikodym derivative is as given above.

Define the product measures $\hat{\mu}, \hat{\nu}$ on \mathbb{R}^{∞} by

$$\hat{\mu} = \prod_{j=1}^{\infty} \hat{\mu}_j, \quad \hat{\nu} = \prod_{j=1}^{\infty} \hat{\nu}_j,$$

where $\hat{\mu}_j = N(0, \lambda_j)$, $\hat{\nu}_j = N(0, \gamma_j)$. As a consequence of a result of Kakutani, see [Prato and Zabczyk \(2002\)](#) Proposition 1.3.5, we have that $\hat{\mu} \sim \hat{\nu}$ with

$$\begin{aligned} \frac{d\hat{\mu}}{d\hat{\nu}}(x) &= \prod_{j=1}^{\infty} \frac{d\hat{\mu}_j}{d\hat{\nu}_j}(x_j) \\ &= \prod_{j=1}^{\infty} \frac{\gamma_j}{\lambda_j} \cdot \exp\left(\frac{1}{2} \sum_{j=1}^{\infty} \left(\frac{1}{\gamma_j} - \frac{1}{\lambda_j}\right) x_j^2\right). \end{aligned}$$

We associate H with \mathbb{R}^{∞} via the map $G : H \rightarrow \mathbb{R}^{\infty}$, given by

$$G_j u = \langle u, \varphi_j \rangle, \quad j \geq 1.$$

Note that the image of G is $\ell^2 \subseteq \mathbb{R}^{\infty}$, and $G : H \rightarrow \ell^2$ is an isomorphism. Since A and B are trace-class, samples from $\hat{\mu}$ and $\hat{\nu}$ almost surely take values in ℓ^2 . G^{-1} is hence almost surely defined on samples from $\hat{\mu}$ and $\hat{\nu}$. Define the translation map $T_m : H \rightarrow H$ by $T_m u = u + m$. Then by the Karhunen–Loève theorem, the measures μ and ν can be expressed as the push forwards

$$\mu = T_m^{\#}(G^{-1})^{\#} \hat{\mu}, \quad \nu = T_m^{\#}(G^{-1})^{\#} \hat{\nu}.$$

Now let $f : H \rightarrow \mathbb{R}$ be bounded measurable, then we have

$$\begin{aligned} \int_H f(u) \mu(du) &= \int_H f(u) [T_m^{\#}(G^{-1})^{\#} \hat{\mu}](du) \\ &= \int_{\mathbb{R}^{\infty}} f(G^{-1}x + m) \hat{\mu}(dx) \\ &= \int_{\mathbb{R}^{\infty}} f(G^{-1}x + m) \frac{d\hat{\mu}}{d\hat{\nu}}(x) \hat{\nu}(dx) \\ &= \int_H f(u) \frac{d\hat{\mu}}{d\hat{\nu}}(G(u - m)) [T_m^{\#}(G^{-1})^{\#} \hat{\nu}](du) \\ &= \int_H f(u) \frac{d\hat{\mu}}{d\hat{\nu}}(G(u - m)) \nu(du). \end{aligned}$$

From this it follows that we have

$$\begin{aligned} \frac{d\mu}{d\nu}(u) &= \frac{d\hat{\mu}}{d\hat{\nu}}(G(u - m)) \\ &= \prod_{j=1}^{\infty} \frac{\gamma_j}{\lambda_j} \cdot \exp\left(\frac{1}{2} \sum_{j=1}^{\infty} \left(\frac{1}{\gamma_j} - \frac{1}{\lambda_j}\right) \langle u - m, \varphi_j \rangle^2\right). \end{aligned}$$

□

Remark 5 The proposition above, in the case $m = 0$, is given as Theorem 1.3.7 in [Prato and Zabczyk \(2002\)](#) except that, there, the factor before the exponential is omitted. This is because it does not depend on u , and all measures involved are probability measures and hence normalized. We retain the factor as we are interested in the precise value of the derivative for the MCMC algorithm, in particular its dependence on the length-scale.

References

- Adler, A., Lionheart, W.R.B.: Uses and abuses of EIDORS: an extensible software base for EIT. *Physiol. Meas.* **27**(5), S25–S42 (2006)
- Agapiou, S., Bardsley, J.M., Paspaliopoulos, O., Stuart, A.M.: Analysis of the Gibbs sampler for hierarchical inverse problems. *J. Uncertain. Quantif.* **2**, 511–544 (2014)

- Agapiou, S., Bardsley, J.M., Papaspiliopoulos, O., Stuart, A.M.: Analysis of the Gibbs sampler for hierarchical inverse problems. *SIAM/ASA J. Uncertain. Quantif.* **2**(1), 511–544 (2014)
- Alvarez, L., Morel, J.M.: Formalization and computational aspects of image analysis. *Acta Numer.* **3**, 1–59 (1994)
- Arbogast, T., Wheeler, M.F., Yotov, I.: Mixed finite elements for elliptic problems with tensor coefficients as cell-centered finite differences. *SIAM J. Numer. Anal.* **34**, 828–852 (1997)
- Bear, J.: *Dynamics of Fluids in Porous Media*. Dover Publications, New York (1972)
- Beskos, A., Roberts, G.O., Stuart, A.M., Voss, J.: MCMC methods for diffusion bridges. *Stoch. Dyn.* **8**, 319–350 (2008)
- Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer, New York (2006)
- Bolin, D., Lindgren, F.: Excursion and contour uncertainty regions for latent Gaussian models. *J. R. Stat. Soc. Ser. B* **77**(1), 85–106 (2015)
- Borcea, L.: Electrical impedance tomography. *Inverse Probl.* **18**, R99–R136 (2002)
- Burger, M.: A level set method for inverse problems. *Inverse Probl.* **17**(5), 1327–1355 (2001)
- Calvetti, D., Somersalo, E.: A Gaussian hypermodel to recover blocky objects. *Inverse Probl.* **23**(2), 733–754 (2007)
- Calvetti, D., Somersalo, E.: Hypermodels in the Bayesian imaging framework. *Inverse Probl.* **24**(3), 34013 (2008)
- Carrera, J., Neuman, S.P.: Estimation of aquifer parameters under transient and steady state conditions: 3. application to synthetic and field data. *Water Resour. Res.* **22**(2), 228–242 (1986)
- Chung, E.T., Chan, T.F., Tai, X.-C.: Electrical impedance tomography using level set representation and total variational regularization. *J. Comput. Phys.* **205**(1), 357–372 (2005)
- Cotter, S.L., Roberts, G.O., Stuart, A.M., White, D.: MCMC methods for functions modifying old algorithms to make them faster. *Stat. Sci.* **28**(3), 424–446 (2013)
- Da Prato, G., Zabczyk, J.: *Second Order Partial Differential Equations in Hilbert Spaces*, vol. 293. Cambridge University Press, Cambridge (2002)
- Dashti, M., Stuart, A.M.: The Bayesian approach to inverse problem. In: Ghanem, R., Higdon, D., Owhadi, H. (eds.) *Handbook of Uncertainty Quantification*. Springer, Heidelberg (2016)
- Dorn, O., Lesselier, D.: Level set methods for inverse scattering. *Inverse Probl.* **22**(4), R67–R131 (2006)
- Dunlop, M.M., Stuart, A.M.: The Bayesian formulation of EIT: analysis and algorithms. [arXiv:1508.04106](https://arxiv.org/abs/1508.04106) (2015)
- Filippone, M., Girolami, M.: Pseudo-marginal Bayesian inference for Gaussian processes. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(11), 2214–2226 (2014)
- Franklin, J.N.: Well posed stochastic extensions of ill posed linear problems. *J. Math. Anal. Appl.* **31**(3), 682–716 (1970)
- Fuglstad, G.-A., Simpson, D., Lindgren, F., Rue, H.: Interpretable priors for hyperparameters for Gaussian random fields. [arXiv:1503.00256](https://arxiv.org/abs/1503.00256) (2015)
- Geirsson, Ó.P., Hrafnkelsson, B., Simpson, D., Siguroarson H.: The MCMC split sampler: a block Gibbs sampling scheme for latent Gaussian models. [arXiv:1506.06285](https://arxiv.org/abs/1506.06285) (2015)
- Girolami, M., Calderhead, B.: Riemann manifold Langevin and Hamiltonian Monte Carlo methods. *J. R. Stat. Soc. Ser. B* **73**(2), 123–214 (2011)
- Hairer, M., Stuart, A.M., Vollmer, S.J.: Spectral gaps for Metropolis–Hastings algorithms in infinite dimensions. *Ann. Appl. Prob.* **24**, 2455–2490 (2014)
- Hanke, M.: A regularizing Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems. *Inverse Probl.* **13**, 79–95 (1997)
- Iglesias, M.A.: A regularizing iterative ensemble Kalman method for PDE-constrained inverse problems. *Inverse Probl.* **32**(2), 025002 (2016)
- Iglesias, M.A., Dawson, C.: The representer method for state and parameter estimation in single-phase Darcy flow. *Comput. Methods Appl. Mech. Eng.* **196**(1), 4577–4596 (2007)
- Iglesias, M.A., Law, K.J.H., Stuart, A.M.: The ensemble Kalman filter for inverse problems. *Inverse Probl.* **29**(4), 045001 (2013)
- Iglesias, M.A., Lu, Y., Stuart, A.M.: A Bayesian level set method for geometric inverse problems. *Interfaces and Free Boundary Problems*, (2016) (to appear)
- Kaipio, J.P., Somersalo, E.: *Statistical and Computational Inverse Problems*. Springer, New York (2005)
- Lasanen, S.: Non-Gaussian statistical inverse problems. Part I: posterior distributions. *Inverse Probl. Imagin.* **6**(2), 215–266 (2012)
- Lasanen, S.: Non-Gaussian statistical inverse problems. Part II: posterior convergence for approximated unknowns. *Inverse Probl. Imag.* **6**(2), 215–266 (2012)
- Lasanen, S., Huttunen, J.M.J., Roininen, L.: Whittle–Matérn priors for Bayesian statistical inversion with applications in electrical impedance tomography. *Inverse Probl. Imag.* **8**(2), 561–586 (2014)
- Lehtinen, M.S., Paivarinta, L., Somersalo, E.: Linear inverse problems for generalised random variables. *Inverse Probl.* **5**(4), 599–612 (1999)
- Lindgren, F., Rue, H.: Bayesian spatial modelling with R-INLA. *J. Stat. Softw.* **63**(19), 63–76 (2015)
- Lorentzen, R.J., Flornes, K.M., Naevdal, G.: History matching channelized reservoirs using the ensemble Kalman filter. *Soc. Pet. Eng. J.* **17**(1), 122–136 (2012)
- Lorentzen, R.J., Nævdal, G., Shafieirad, A.: Estimating facies fields by use of the ensemble Kalman filter and distance functions—applied to shallow-marine environments. *Soc. Pet. Eng. J.* **3**, 146–158 (2012)
- Mandelbaum, A.: Linear estimators and measurable linear transformations on a Hilbert space. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **65**(3), 385–397 (1984)
- Matérn, B.: *Spatial Variation*, vol. 36. Springer Science & Business Media, Berlin (2013)
- Marshall, R.J., Mardia, K.V.: Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**(1), 135–146 (1984)
- Osher, S., Sethian, J.A.: Fronts propagating with curvature dependent speed: algorithms based on Hamilton–Jacobi formulations. *J. Comput. Phys.* **79**, 12–49 (1988)
- Ping, J., Zhang, D.: History matching of channelized reservoirs with vector-based level-set parameterization. *Soc. Pet. Eng. J.* **19**, 514–529 (2014)
- Rasmussen, C.E., Williams, C.K.I.: *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge (2006)
- Robert, C., Casella, G.: *Monte Carlo Statistical Methods*. Springer Science & Business Media, Berlin (2013)
- Santosa, F.: A level-set approach for inverse problems involving obstacles. *ESAIM I*(1), 17–33 (1996)
- Sapiro, G.: *Geometric Partial Differential Equations and Image Analysis*. Cambridge University Press, Cambridge (2006)
- Srinivasa Varadhan, S.R.: *Probability Theory*. Courant Lecture Notes. Courant Institute of Mathematical Sciences, New York (2001)
- Somersalo, E., Cheney, M., Isaacson, D.: Existence and uniqueness for electrode models for electric current computed tomography. *SIAM J. Appl. Math.* **52**(4), 1023–1040 (1992)
- Stein, M.L.: *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Science & Business Media, Berlin (2012)
- Stuart, A.M.: Inverse problems : a Bayesian perspective. *Acta Numer.* **19**, 451–559 (2010)
- Tai, X.-C., Chan, T.F.: A survey on multiple level set methods for identifying piecewise constant functions. *Int. J. Numer. Anal. Model.* **1**(1), 25–48 (2004)
- Tierney, L.: A note on Metropolis–Hastings kernels for general state spaces. *Ann. Appl. Prob.* **8**(1), 1–9 (1998)

-
- van der Vaart, A.W., van Zanten, J.H.: Adaptive Bayesian estimation using a Gaussian random field with inverse gamma bandwidth. *Ann. Stat.* **37**, 2655–2675 (2009)
- Xie, J., Efendiev, Y., Datta-Gupta, A.: Uncertainty quantification in history matching of channelized reservoirs using Markov chain level set approaches. *Soc. Pet. Eng.* **1**, 49–76 (2011)
- Zhang, H.: Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Am. Stat. Assoc.* **99**(465), 250–261 (2004)