Journal of Computational Physics 228 (2009) 6184–6211

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

A multi-resolution, non-parametric, Bayesian framework for identification of spatially-varying model parameters

P.S. Koutsourelakis *

School of Civil and Environmental Engineering, Cornell University, Ithaca, NY 14853, USA

ARTICLE INFO

Article history:
Received 3 October 2008
Received in revised form 7 May 2009
Accepted 11 May 2009
Available online 19 May 2009

Keywords:
Uncertainty
Complex systems
Statistical learning
Monte Carlo
Bayesian

ABSTRACT

This paper proposes a hierarchical, multi-resolution framework for the identification of model parameters and their spatially variability from noisy measurements of the response or output. Such parameters are frequently encountered in PDE-based models and correspond to quantities such as density or pressure fields, elasto-plastic moduli and internal variables in solid mechanics, conductivity fields in heat diffusion problems, permeability fields in fluid flow through porous media etc. The proposed model has all the advantages of traditional Bayesian formulations such as the ability to produce measures of confidence for the inferences made and providing not only predictive estimates but also quantitative measures of the predictive uncertainty. In contrast to existing approaches it utilizes a parsimonious, non-parametric formulation that favors sparse representations and whose complexity can be determined from the data. The proposed framework in non-intrusive and makes use of a sequence of forward solvers operating at various resolutions. As a result, inexpensive, coarse solvers are used to identify the most salient features of the unknown field(s) which are subsequently enriched by invoking solvers operating at finer resolutions. This leads to significant computational savings particularly in problems involving computationally demanding forward models but also improvements in accuracy. It is based on a novel, adaptive scheme based on Sequential Monte Carlo sampling which is embarrassingly parallelizable and circumvents issues with slow mixing encountered in Markov Chain Monte Carlo schemes. The capabilities of the proposed methodology are illustrated in problems from nonlinear solid mechanics with special attention to cases where the data is contaminated with random noise and the scale of variability of the unknown field is smaller than the scale of the grid where observations are collected.

1. Introduction

The prodigious advances in computational modeling of physical processes and the development of highly non-linear, multiscale and multiphysics models poses several challenges in parameter identification. We are frequently using large, forward models which imply a significant computational burden, in order to analyze complex phenomena.The extensive use of such models poses several challenges in parameter identification as the accuracy of the results provided depends strongly on assigning proper values to the various model parameters. In mechanics of materials, accurate mechanical property identification can guide damage detection and an informed assessment of the system's reliability [36]. Identifying property-cross

* Tel.: +1 607 254 5441.
E-mail address: pk285@cornell.edu

0021-9991/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2009.05.016

correlations can lead to the design of multi-functional materials [61]. In biomechanics, the detection of variations in mechanical properties of human tissue can reveal the appearance of diseases (arteriosclerosis, malignant tumors) but can also be used to assess the effectivity of various treatments [4,20]. Permeability estimation for soil transport processes can assist in detection of contaminants, oil exploration etc [22,68].

We consider phenomena described by a set of (coupled) elliptic, parabolic or hyperbolic PDEs and associated boundary (and initial) conditions:

$$\mathcal{A}(\boldsymbol{y}(\boldsymbol{x}); f(\boldsymbol{x})) = 0, \quad \forall x \in \mathcal{D} \tag{1}$$

where $\mathcal{A}$ denotes the differential operator defined on a domain $\mathcal{D} \in \mathbb{R}^d$, where d is the number of spatial dimensions. $\mathcal{A}$ depends on spatially varying coefficients $f(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{D}$. Advances in computational mathematics have given rise to several efficient solvers for a wide-range of such systems and have revolutionized simulation-based analysis and design [52]. Our primary interest is to identify $f(\boldsymbol{x})$ from a set of (potentially noisy) measurements of the response $\boldsymbol{y}_i = \boldsymbol{y}(\boldsymbol{x}_i)$ at a number of distinct locations $\boldsymbol{x}_i \in \mathcal{D}$. In the case of time-dependent PDEs, the available data might also be indexed by time. Several different processes in solid and fluid mechanics, transport phenomena, heat diffusion etc fall under this general setting and even though the coefficients $f(\boldsymbol{x})$ have a different physical interpretation, the associated inverse problems exhibit similar mathematical characteristics.

Two basic approaches have been followed in addressing problems of data-driven parametric identification. On one hand, deterministic optimization techniques which attempt to minimize the sum of the squares of the deviations between model predictions and observations. Gradient or global, intrusive or non-intrusive techniques are introduced for performing the optimization task. Usually the objective function is augmented with regularization terms (e.g. Tikhonov regularization [58] which alleviate issues with the ill-posednesss of the problem [5,18,26,37,59,64]. Such deterministic inverse techniques based on exact matching or least-squares optimization, lead to point estimates of unknowns without rigorously considering the statistical nature of system uncertainties and without providing quantification of the uncertainty in the inverse solution.

The direct stochastic counterpart of optimization methods involves frequentist approaches based on maximum likelihood estimators that aim at maximizing the probability of observations given the inverse solution maximum [17,19]. In recent years significant attention has been directed towards statistical approaches based on the Bayesian paradigm which attempt to calculate a (posterior) probability distribution function on the parameters of interest. Bayesian formulations offer several advantages as they provide a unified framework for dealing with the uncertainty introduced by the incomplete and noisy measurements and assessing quantitatively resulting inferential uncertainties. Significant successes have been noted in applications such as medical tomography [69], geological tomography [2,24], hydrology [43], petroleum engineering [8,27], as well as a host of other physical, biological, or social systems [41,47,56,67].

Identification of spatially varying model parameters poses several modeling and computational issues. Representations of the parametric fields in existing approaches artificially impose a minimum length scale of variability usually determined by the discretization size of the governing PDEs [43]. Furthermore, they are associated with a very large vector of unknowns. Inference in high-dimensional spaces using standard optimization or sampling schemes (e.g. Markov Chain Monte Carlo (MCMC)), is generally impractical as it requires an exuberant number of calls to the forward simulator in order to achieve convergence. Particularly in Bayesian formulations where the inference results are much richer and involve a distribution rather than a single value for the parameters of interest, the computational effort implied by repeated calls to the forward solver can be enormous and constitute the method impractical for realistic applications. These problems are amplified if the posterior distribution is multi-modal i.e. several significantly different scenaria are likely given the available data. While it is apparent that, computationally inexpensive, coarser scale simulations can assist the identification process [13], the critical task of efficiently transferring the information across resolutions still remains [30,49,68]. Previous attempts using parallel tempering (e.g. [32] or hierarchical representations based on Markov trees [65] require performing inference on representations at various resolutions simultaneously.

In the present paper we adopt a nonparametric model which is independent of the grid of the forward solver and is reminiscent of non-parametric kernel regression methods. The unknown parametric field is approximated by a superposition of kernel-type functions centered at various locations. The cardinality of the representation, i.e. the number of such kernels, is treated as an unknown to be inferred in the Bayesian formulation. This gives rise to a very flexible model that is able to adapt to the problem and the data at hand and find succinct representations of the parametric field of interest. Prior information on the scale of variability can be directly introduced in the model.

Inference is performed using Sequential Monte Carlo samplers. They utilize a set of random samples, named particles, which are propagated using simple importance sampling, resampling and updating/rejuvenation mechanisms. The algorithm is directly parallelizable as the evolution of each particle is by-and-large independent of the rest. The sequence of distributions defined is based on using solvers that operate on different resolutions and which successively produce finer discretizations. This results in an efficient hierarchical approach that makes use of the results from solvers operating at the coarser scales in order to update them based on analyses on a finer scale. The particulate approximations produced provide concise representations of the posterior which can be readily updated if more data become available or if more accurate solvers are employed.

2. Methodology

2.1. Hierarchical bayesian model

The central goal of this work is to build mathematical methods that utilize limited and noisy observations/measurements in order to identify the spatial variability of model parameters. Given the significant uncertainty arising from the random noise, lack of data and model error, point estimates are of little use. Furthermore it is important to quantify the confidence in the estimates made but also in the predictive ability of the model of interest. To that end we adopt a Bayesian perspective. Bayesian formulations differ from classical statistical approaches (frequentist) in that all unknown parameters (denoted by $\boldsymbol{\theta}$) are treated as random. Hence the results of the inference process are not point estimates but distribution functions.

The basic elements of Bayesian models are the *likelihood* function $L(\boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{\theta})$ which is a conditional probability distribution and gives a (relative) measure of the propensity of observing data $\boldsymbol{y}$ for a given model configuration specified by the parameters $\boldsymbol{\theta}$. The likelihood function is also encountered in frequentist formulations where the unknown model parameters $\boldsymbol{\theta}$ are determined by maximizing $L(\boldsymbol{\theta})$. This could be thought as the probabilistic equivalent of deterministic optimization techniques commonly used in inverse problems. It can suffer from the same issues related to the ill-posedeness of the problem. The second component of Bayesian formulations is the *prior* distribution $p(\boldsymbol{\theta})$ which encapsulates in a probabilistic manner any knowledge/information/insight that is available to the analyst prior to observing the data. Although the prior is a point of frequent criticism due to its inherently subjective nature, it can prove extremely useful in engineering contexts as it provides a mathematically consistent vehicle for injecting the analyst's insight and physical understanding. The combination of *prior* and *likelihood* based on Bayes' rule yields the *posterior* distribution $\pi(\boldsymbol{\theta})$ which probabilistically summarizes the information extracted from the data with regards to the unknown $\boldsymbol{\theta}$:

$$\pi(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\boldsymbol{y})} \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{2}$$

Hence Bayesian formulations allow for the possibility of multiple solutions – in fact any θ in the support of the likelihood and the prior is admissible – whose *relative plausibility* is quantified by the posterior. Credible intervals can be readily estimated from the posterior which quantify inferential uncertainties about the unknowns.

Without loss of generality, we postulate the existence of a deterministic, forward model which in most cases of practical interest corresponds to a Finite Element or Finite Difference model of the governing differential equations. Naturally, forward models allow for various levels of discretization of the spatial domain and let r denote the resolution they operate upon (larger r implies finer resolution). In this paper, forward solvers are viewed as *messengers*, that carry information about the underlying material properties as they manifest themselves in the response (mechanical, thermal etc) of the medium of interest. This is especially true in the context of recently developed upscaling schemes [12,16,33,34,39,42,55,62] which attempt to capture the effect of finer scale material variability while operating on a coarser grid. In general, the finer the resolution of the forward solver, the more information this provides. This however comes at the expense of computational effort. It is not unusual that the sufficient resolution of the property fluctuations in many systems of practical interest requires several CPU-hours for a single analysis. Despite the fidelity and accuracy of such high-resolution solvers, they can be of little use in the context of parameter identification as they will generally have to be called upon several times and several system analyses will have to be performed.

Hence an accurate but expensive *messenger* is not the optimal choice if several pieces of information need to be communicated. In many cases however, the fidelity of the message can be compromised if the expense associated with the messenger is smaller. This is especially true if the loss of accuracy can be quantified, measures of confidence can be provided and furthermore if it leads to the same decisions/predictions. In this project we propose a consistent framework for using faster but less-accurate forward solvers operating on coarser resolutions in order to expedite property identification. Furthermore these solvers provide a natural hierarchy of models that if appropriately coupled can further expedite the identification process. Following the analog introduced earlier, we propose using inexpensive messengers (coarse scale solvers), several times to communicate the most pivotal pieces of information and more expensive messengers (fine scale solvers) fewer times to pass on some of the finer details (Fig. 1).

In the remainder of this sub-section, we discuss the basic components of the Bayesian model proposed, with particular emphasis on the prior for the unknown parametric fields. We then present (sub-Section 2.2) the proposed inference techniques for the determination of the posterior.

2.1.1. Likelihood specification

Let $\boldsymbol{F}^r = \{F_i^r\} : \mathcal{G} \to \mathcal{E}$ denote the vector-valued mapping implied by the forward model (operating at resolution r), which given $f(x) \in \mathcal{G}$ (Eq. (1)) provides the values of response quantities represented by the data $\boldsymbol{y} = \{y_i\} \in \mathcal{E}$. This function is the discretized version of the inverse of the differential operator $\mathcal{A}$ in Eq. (1) parameterized by $f(\boldsymbol{x})$. Each evaluation of $\boldsymbol{F}^r$ for a specific field $f(\boldsymbol{x})$ implies a call to the forward solver (e.g. Finite Elements) that operates on a discretization/resolution r. In the proposed framework, the function $\boldsymbol{F}^r$ will be treated as a black box. Naturally data and model predictions will deviate when the former are obtained experimentally due to the unavoidable noise in the measurements. Most importantly perhaps this deviation can be the result of the model not fully capturing the salient physics either because the governing PDEs are an idealization or because of the discretization error in their solution. We postulate the following relationship:

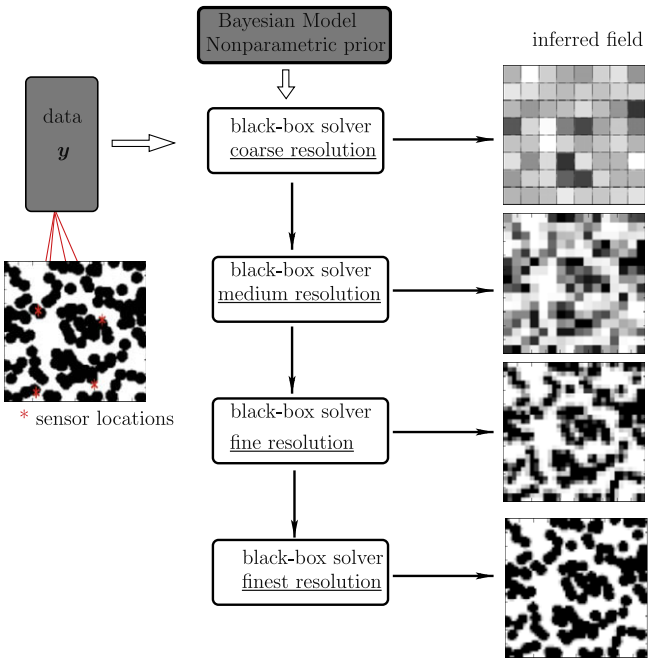


Fig. 1. Hierarchy of solvers operating on different resolutions.

$$\underbrace{y_i}_{\text{datum } i} = \underbrace{F_i^{(r)}(f(\boldsymbol{x}))}_{\text{model prediction}} + e_i^{(r)} \quad i = 1, 2, \ldots, n \tag{3}$$

where $e_i^{(r)}$ quantify the deviation between model predictions and data, and which will naturally depend on the resolution r of the forward solver. Quite frequently the data available to us are in the form of disparate observations, that correspond to different physical phenomena (e.g. temperatures and displacements in a thermo-mechanical problem) in which case the computational model corresponds to a coupled multiphysics solver.

The probabilistic model for e_i^r in Eq. (3) gives rise to the *likelihood function* (Eq. (2)). In the simplest case where $e_i^{(r)}$ are assumed independent, normal variates with zero mean and variance σ_r^2:

$$p_r(y_i|f(x),\sigma_r) \propto \frac{1}{\sigma_r}\exp\left\{-\frac{1}{2}\frac{(y_i - F_i^{(r)}(f(\boldsymbol{x})))^2}{\sigma_r^2}\right\} \quad \text{and} \quad p_r(\boldsymbol{y}|f(x),\sigma_r) \propto \frac{1}{\sigma_r^n}\exp\left\{-\frac{1}{2\sigma_r^2}\sum_{i=1}^{n}(y_i - F_i^{(r)}(f(\boldsymbol{x})))^2\right\} \tag{4}$$

More complex models which can account for the spatial dependence of the error variance σ_r^2 or the detection of events associated with sensor malfunctions at certain locations, can readily be formulated.

In general the variances σ_r^2 are unknown (particularly the component that pertains to model error) and should be inferred from the data. When a conjugate, $\Gamma(a,b)$ prior is adopted for σ_r^{-2}, the error variances can be integrated out from Eq. (4) further simplifying the likelihood:

$$L_r(f(x)) = p(\boldsymbol{y}|f(x)) \propto \frac{\Gamma(a + n/2)}{\left(b + \frac{1}{2}\sum_{i=1}^{n}(y_i - F_i^{(r)}(f(\boldsymbol{x}))^2\right)^{a+n/2}} \tag{5}$$

where $\Gamma(z) = \int_0^{+\infty} t^{z-1}e^{-t}dt$ is the gamma function.

It should be noted that in some works [31,38], explicit distinction between model and observation errors is made, postulating a relation of the following form:

$$observation/data = model\ prediction + model\ error + observation\ error \tag{6}$$

As it has been observed [70], independently of the amount of data available to us, these three components are not *identifiable*, meaning several different values can be equally consistent with the data. This however does not imply that all possible values are equally plausible. For example a large number of values of the observation error that are all positive or all negative (for all observations) are not consistent with the perception of random noise but most likely imply a bias of the model or perhaps a miscalibrated sensors used to collect the data. Bayesian formulations are highly suited for such problems as they provide a natural way of quantifying a priori and a posteriori relative measures of plausibility. In the following we restrict the presentation on models of Eq. (3) as the focus of is on identifying the scale of variability of material properties $f(\boldsymbol{x})$.

2.1.2. Prior specification

The most critical component involves the prior specification for the unknown material properties as represented by $f(x)$. In existing Bayesian [36,67], but also deterministic (optimization-based), formulations, $f(x)$ is discretized according to the spatial resolution of the forward solver. For example, in cases where finite elements are used, the property of interest is assumed constant within each element and therefore the vector of unknowns is of dimension equal to the number of elements. This offers obvious implementation advantages but also poses some difficulties since the scale of variability of material properties is implicitly selected by the solver rather than the data. This is problematic in several ways. On one hand if the scale of variability is larger than the grid, a waste of resources takes place, at the solver level which has to be run at unnecessarily fine resolutions, and at the level of the inference process which is impeded by the unnecessarily large dimension of the vector of unknowns. Furthermore, as the number of unknowns is much larger by comparison to the amount of data it can lead to *overfitting*. This will produce erroneous or even absurd values for the unknowns that may nevertheless fit perfectly the data. Such solutions will have negligible *predictive ability* and would be useless in decision making. On the other hand, if the scale of variability is smaller than the grid, it cannot be identified even if the solver provides sufficient information for discovering this possibility.

In order to increase the flexibility of the model, we base our prior models for the unknown field(s) $f(\boldsymbol{x})$ on the convolution representation of a Gaussian process. An alternative representation of a stationary Gaussian process involves a convolution of a white noise process $a(\boldsymbol{x})$ with a smoothing kernel $K(\cdot;\phi)$ depending on a set of parameters ϕ [3,28]:

$$f(\boldsymbol{x}) = \int K(\boldsymbol{x}-\boldsymbol{z};\phi)a(\boldsymbol{z})d\boldsymbol{z} \tag{7}$$

The kernel form determines essentially the covariance of the resulting process, since:

$$cov(f(\boldsymbol{x}_1,f(\boldsymbol{x}_2)) = E[f(\boldsymbol{x}_1,f(\boldsymbol{x}_2)] = \int K(\boldsymbol{x}_1-\boldsymbol{z};\phi)K(\boldsymbol{x}_2-\boldsymbol{z};\phi)d\boldsymbol{z} \tag{8}$$

For computational purposes, a discretized version of Eq. (7) is used:

$$f(\boldsymbol{x}) = \sum_{j=1}^{k} a(\boldsymbol{z}_j)K(\boldsymbol{x}-\boldsymbol{z}_j;\phi) = \sum_{j=1}^{k} a_j K(\boldsymbol{x}-\boldsymbol{x}_j;\phi) \tag{9}$$

In order to increase the expressive ability of the aforementioned model we introduce two improvements. Firstly we consider that the set of kernel parameters ϕ is spatially varying resulting in a non-stationary process:

$$f(\boldsymbol{x}) = a_0 + \sum_{j=1}^{k} a_j K_j(\boldsymbol{x};\boldsymbol{\phi}_j) \quad x \in D \tag{10}$$

where a_0 corresponds to a value of ϕ_0 such that the corresponding kernel is 1 everywhere. Such representations can be viewed as a radial basis network as in [60,63]. Furthermore by interpreting the kernels as basis functions, Eq. (10) it can be seen as an extension of the representer theorem of Kimeldorf and Wahba [40]. Overcomplete representations as in Eq. (10) have been advocated because they have greater robustness in the presence of noise, can be sparser, and can have greater flexibility in matching structure in the data [45,46]. One possible selection for the functional form of K_j, that also has an intuitive parameterization with regards to the scale of variability in the material properties, is isotropic, Gaussian kernels:

$$K(\boldsymbol{x};\boldsymbol{\phi}_j = (\boldsymbol{x}_j,\tau_j)) = \exp\{-\tau_j\|\boldsymbol{x}-\boldsymbol{x}_j\|^2\} \tag{11}$$

The parameters τ_j directly correspond to the scale of variability of $f(\boldsymbol{x})$. Large τ_j's imply narrowly concentrated fluctuations and large values slower varying fields. The center of each kernel is specified by the location parameter $\boldsymbol{x}_j$. Other functional forms (e.g. discontinuous) can also be used on their own or in combinations to enrich the expressivity of the expansion in Eq. (10). Wavelets, steerable wavelets, segmented wavelets, Gabor dictionaries, multiscale Gabor dictionaries, wavelet packets, cosine packets, chirplets, warplets, and a wide range of other dictionaries that have been developed in various contexts [6] offer several possibilities.

The second important improvement is that we allow the size of the expansion k to vary. It is obvious that such an assumption is consistent with the *principle of parsimony*, which states that prior models should make as few assumptions as possible and allow their complexity to be inferred from the data.

Hence the *cardinality*of the model, i.e. the number of basis functions k is the key unknown that must be determined so as to provide a good interpretation of the observables.

Independently of the form of the kernel adopted, the important, common characteristic of all such approximations (as in Eq. (10)) is that the field representation *does not depend on the resolution of the forward model*. The latter affects inference only through the black-box functions F_i^r (Eq. (3), Fig. 1)) as it will be illustrated in the next sections.

The parameters of the prior model adopted consist of:

- k: the number of kernel functions needed,
- $\{a_j\}_{j=1}^{k}$, the coefficients of the expansion in Eq. (10). Each of those can be a scalar or vector depending on the number of material property fields we want to infer simultaneously. For example, in a problem of thermo-mechanical coupling

where the data consists of temperatures and displacements and we want to identify elastic modulus and conductivity, each a_j will be a vector in $\mathbb{R}^2$.

- $\{\tau_j\}_{j=1}^k$ the precision parameters of each kernel which pertain to the scale of the unknown field(s), and
- $\{\boldsymbol{x}_j\}_{j=1}^k$ the locations of the kernels which are points in $\mathcal{D}$.

In accordance with the Bayesian paradigm, all unknowns are considered random and are assigned prior distributions which quantify any information, knowledge, physical insight, mathematical constraints that is available to the analyst before the data is processed. Naturally, if specific prior information is available it can be reflected on the prior distributions. We consider prior distributions of the following form (excluding hyperparameters):

$$p(k,\{a_j\}_{j=0}^k,\{\tau_j\}_{j=1}^k,\{\boldsymbol{x}_j\}_{j=1}^k) \propto p(k)\times p(\{a_j\}_{j=0}^k|k)\times p(\{\tau_j\}_{j=1}^k|k)\times p(\{\boldsymbol{x}_j\}_{j=1}^k)) \tag{12}$$

In order to increase the robustness of the model and exploit structural dependence we adopt a hierarchical prior model [23].

2.1.2.1. Model size. Pivotal to the robustness and expressivity of the model is the selection of the model size, i.e. of the number of kernel functions k in Eq. (10). This number is unknown a priori and in the absence of specific information, *sparse* representations should be favored. This is not only advantageous for computational purposes, as the number of unknown parameters is proportional to k, but also consistent with the parsimony of explanation principle or Occam's razor [35,51,53]. For that purpose, we propose a truncated Poisson prior for k:

$$p(k|\lambda) \propto \begin{cases} e^{-\lambda}\frac{\lambda^k}{k!} & \text{if } k \leqslant k_{max} \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

The truncation parameter k_{max} is selected based on computer memory limitations and defines the support of the prior. This prior allows for representations of various cardinalities to be assessed simultaneously with respect to the data. As a result the number of unknowns is not fixed and the corresponding posterior has support on spaces of different dimensions as discussed in more detail in the sequence. In this work, an exponential hyper-prior is used for the hyper-parameter λ to allow for greater flexibility and robustness i.e. $p(\lambda|s) = s\exp\{-\lambda s\}$. After integrating out λ we obtain:

$$p(k|s) \propto \frac{1}{(s+1)^{k+1}}, \quad \text{for} \quad k = 0,1,\ldots,k_{max} \tag{14}$$

2.1.2.2. Scale. The most critical perhaps parameters of the model are $\{\tau_j\}_{j=1}^k$ which control the *scale of variability* in the approximation of the unknown field(s). If prior information about this is available then it can be readily accounted for by appropriate prior specification. In the absence of such information however multiple possibilities exist. In contrast to deterministic optimization techniques where ad-hoc *regularization* assumptions are made, in the Bayesian framework proposed possible solutions are evaluated with respect to their *plausibility* as quantified by the posterior distribution. This provides a unified interpretation of various assumptions that are made regarding the priors of the parameters involved. For example, consider a general $\Gamma(a_\tau,b_\tau)$ prior:

$$p(\{\tau_j\}_{j=1}^k|k,a_\tau,b_\tau) = \prod_{j=1}^k \frac{b_\tau^{a_\tau}}{\Gamma(a_\tau)}\tau_j^{a_\tau-1}\exp(-b_\tau\tau_j) \tag{15}$$

This has a mean a_τ/b_τ and coefficient of variation $1/\sqrt{a_\tau}$. Diffuse versions can be adopted by selecting small a_τ. A non-informative prior $p(\tau_j)\propto 1/\tau_j$ arises as a special case for $a_\tau = 2$ and $b_\tau = 0$ which is invariant under rescaling. Furthermore. it offers an interesting physical interpretation as it favors "slower" varying representations (i.e. smaller τ's). In order to automatically determine the mean of the Gamma prior, we express $b_\tau = \mu_j a_\tau$ where μ_j is a location parameter for which an Exponential hyper-prior is used with a hyper-parameter a_μ i.e. $p(\mu_j) = \frac{1}{a_\mu}e^{-\mu_j/a_\mu}$. Integrating out the μ_j's leads to following prior:

$$p(\{\tau_j\}_{j=1}^k|k,a_\tau,a_\mu) = \prod_{j=1}^k \frac{\Gamma(a_\tau+1)}{\Gamma(a_\tau)}a_\tau^{a_\tau}\frac{\tau_j^{(a_\tau-1)}}{a_\mu}\frac{1}{(a_\tau\tau_j+a_\mu^{-1})^{(a_\tau+1)}} \tag{16}$$

2.1.2.3. Other parameters. For the coefficients a_j a multivariate normal prior was adopted:

$$\{a_j\}_{j=0}^k|k,\sigma_a^2 \sim N(0,\sigma_a^2\boldsymbol{I}_{k+1}) \tag{17}$$

where $\boldsymbol{I}_{k+1}$ is the $(k+1)\times(k+1)$ identity matrix. The hyper-parameter σ_a^{-2} which controls the spread of the prior is modeled with a gamma distribution $\Gamma(a_0,b_0)$. It can readily be marginalized leading to the following prior for a_j's:

$$p(\{a_j\}_{j=0}^k|k,a_0,b_0) = \frac{1}{(2\pi)^{(k+1)/2}}\frac{\Gamma(a_0+\frac{k+1}{2})}{\left(b_0+\frac{1}{2}\sum_{j=0}^k a_j^2\right)^{a_0+(k+1)/2}} \tag{18}$$

Finally, for the unknown kernel locations $\{\boldsymbol{x}_j\}_{j=1}^k$, a uniform prior in $\mathcal{D}$ is proposed i.e.:

$$p(\{\boldsymbol{x}_j\}_{j=1}^k|k) = \frac{1}{|\mathcal{D}|^k} \tag{19}$$

where $|\mathcal{D}|$ is the length or area or volume of $\mathcal{D}$ in one, two or three dimensions, respectively. Naturally, if prior information is available about subregions with significant property variations this can be incorporated in the prior.

2.1.2.4. Complete model. Let $\boldsymbol{\theta}_k = \{\{a_j\}_{j=0}^k, \{\tau_j\}_{j=1}^k, \{\boldsymbol{x}_j\}_{j=1}^k\} \in \boldsymbol{\Theta}_k$ denote the vector containing all the unknown parameters and $\boldsymbol{\theta} = (k, \boldsymbol{\theta}_k)$. Since k is also assumed unknown and allowed to vary, the dimension of $\boldsymbol{\theta}_k$ is variable as well and $\boldsymbol{\Theta}_k \triangleq (\mathbb{R}^{k+1} \times (\mathbb{R}^+)^k \times \mathcal{D}^k$.

In 2D for example and assuming a scalar unknown field $f(x)$ in the expansion of Eq. (10) the dimension of $\boldsymbol{\theta}_k$ is $(k+1)+k+2k = 2+4k$. Based on Eq. (12) and Eqs. (14), (15), (18) and (19), the complete prior model is given by:

$$p(\boldsymbol{\theta}|s, a_\tau, a_\mu, a_0, b_0) = \frac{1}{(s+1)^{k+1}} \times \prod_{j=1}^{k} \frac{\Gamma(a_\tau+1)}{\Gamma(a_\tau)} \frac{a_\tau^{a_\tau}}{\tau_j^{(a_\tau-1)}} \frac{1}{a_\mu} \frac{1}{(a_\tau\tau_j + a_\mu^{-1})^{(a_\tau+1)}} \times \frac{1}{(2\pi)^{(k+1)/2}} \frac{\Gamma(a_0+\frac{k+1}{2})}{\left(b_0 + \frac{1}{2}\sum_{j=0}^k a_j^2\right)^{a_0+(k+1)/2}} \times \frac{1}{|\mathcal{D}|^k} \tag{20}$$

The combination of the prior $p(\boldsymbol{\theta})$ with the likelihood $L_r(\boldsymbol{\theta})$ (Eq. (5)) corresponding to a forward solver operating on resolution r, give rise to the *posterior* density $\pi_r(\boldsymbol{\theta})$ which is proportional to:

$$\pi_r(\boldsymbol{\theta}) = p_r(\boldsymbol{\theta}|\boldsymbol{y}) \propto L_r(\boldsymbol{\theta}) p(\boldsymbol{\theta}) \tag{21}$$

It should be noted that the fact that conjugate (hyper-)priors have been adopted for σ_r^2 (Eq. (5)) and the hyperparameters λ (Eq. (14)), b_τ (Eq. (16)) and σ_a^2 (Eq. (18)) allows for their marginalization and gives rise simpler expressions but does not (significantly) affect the computational effort involved. In the case of non-conjugate prior models, MCMC updates for these parameters would have to be performed but those would not require re-calculation of the response $F_i^{(r)}(\cdot)$ (Eq. (5)) which is the most expensive part, as $F_i^{(r)}(\cdot)$ depend on these parameters indirectly (i.e. through $k, a_j, \tau_j, \boldsymbol{x}_j$).

Even though several parameters have been removed from the vector of unknowns $\boldsymbol{\theta}$ and marginalized in the pertinent expressions, the corresponding posteriors can be readily be obtained, or rather be sampled from, once the posteriors $\pi_r(\boldsymbol{\theta})$ have been determined. As it is shown in the numerical examples, of interest could be the variance σ_r^2 of the error term (Eqs. (3) and (4)) which quantifies the magnitude of the deviation between model and data and can serve as a validation metric (in the absence of observation error) or be used for predictive purposes (see Section 2.3). From Eq. (3) and the conjugate prior model adopted for σ_r^2, it can readily be shown that the conditional posterior is given by a Gamma distribution:

$$p(\sigma_r^{-2}, \boldsymbol{\theta}|\boldsymbol{y}) = p(\sigma_r^{-2}|\boldsymbol{\theta})\pi_r(\boldsymbol{\theta}|\boldsymbol{y}) \quad \text{and} \quad p(\sigma_r^{-2}|\boldsymbol{\theta}) = \Gamma\left(a + \frac{n}{2}, b + \frac{\sum_{i=1}^n (y_i - F_i^{(r)}(\boldsymbol{\theta}))^2}{2}\right) \tag{22}$$

In the context of Monte Carlo simulation, this trivially implies that once samples $\boldsymbol{\theta}$ from π_r have been obtained, the samples of σ_r^{-2} can also be drawn from the aforementioned Gamma.

The support of the posteriors π_r lies on $\cup_{k=0}^{k_{max}}\{k\} \times \boldsymbol{\Theta}_k$. Two important points are worth emphasizing. Firstly, Eq. (21) defines a *sequence of posterior densities*, each corresponding to a different likelihood and a different forward solver of resolution r. It is clear that the black-box functions $\boldsymbol{F}^{(r)}$ appearing in the likelihood in Eq. (4) imply *denser* mappings for smaller r. This is because solvers corresponding to coarser resolutions of the governing PDEs are more myopic (compared to solvers at finer resolutions) to small scale fluctuations of the spatially varying model parameters $f(\boldsymbol{x})$ (parameterized by $\boldsymbol{\theta}$). As a result the likelihood functions L_r and the associated posteriors π_r will be flatter and have fewer modes for smaller r. The task of identifying these posteriors becomes increasingly more difficult as we move to solvers of higher refinement (i.e. larger r). It is this feature that we propose of exploiting in the next section in order to increase the accuracy and improve on the efficiency of the inference process. In addition, the posteriors π_r are only known up to a normalizing constant (determining $p(\boldsymbol{y})$ in Eq. (2) involves an infeasible and unnecessary integration in a very high dimensional space). Each evaluation of π_r for a particular $\boldsymbol{\theta}$ requires calculating $\boldsymbol{F}^{(r)}$ and therefore calling the corresponding black-box solver. As each of these runs of the forward solver may involve the solution of very large systems of equations they can be extremely time consuming. It is important therefore to determine π_r not only accurately, but also with the least possible number of calls to the forward solver. Since solvers corresponding to coarser resolutions (smaller r) are faster, it would be desirable to utilize the information they provide in order to reduce the number of calls to more expensive, finer resolution solvers.

2.2. Determining the posterior – inference

The posterior defined above is analytically intractable. For that reason, *Monte Carlo* methods provide essentially the only accurate way to infer π_r. Traditionally *Markov Chain Monte Carlo* techniques (MCMC) have been employed to carry out this task [21,29,43,44,66]. These are based on building a Markov chain that asymptotically converges to the target density (in this case π_r) by appropriately defining a transition kernel. While convergence can be assured under weak conditions [48,54], the rate of convergence can be extremely slow and require a lot of likelihood evaluations and calls to the black-box solver. Par-

ticularly in cases where the target posterior can have multiple modes, very large *mixing times* might be required which constitute the method impractical or infeasible. In addition, MCMC is not directly parallelizable, unless multiple independent chains are run simultaneously and it can be difficult to design a good proposal distribution when operating in high dimensional spaces. More importantly perhaps, standard MCMC is not capable of providing a *hierarchical, multi-resolution* solution to the problem. Consider for example, the case that several samples have been drawn using MCMC from the posterior π_{r_1} corresponding to a solver operating on resolution $r = r_1$. If samples of the posterior π_{r_2} are needed, corresponding to a solver of finer resolution $r_2 > r_1$ but not significantly different from r_1, then MCMC iterations would have to be *initiated anew*. Hence there is no immediate way to exploit the inferences made about π_{r_1} even though the latter might be quite similar to π_{r_2}.

In this work we advocate the use of *Sequential Monte Carlo* techniques (SMC). They represent a set of flexible simulation-based methods for sampling from a sequence of probability distributions [15,50]. As with Markov Chain Monte Carlo methods (MCMC), the target distribution(s) need only be known up to a constant and therefore do not require calculation of the intractable integral in the denominator in Eq. (2). They utilize a set of random samples (commonly referred to as *particles*), which are propagated using a combination of *importance sampling*, *resampling* and MCMC-based *rejuvenation* mechanisms [11,10]. Each of these particles, which can be thought of as a possible configuration of the system's state, is associated with an *importance weight* which is proportional to the posterior value of the respective particle. These weights are updated sequentially along with the particle locations. Hence if $\{\boldsymbol{\theta}_r^{(i)}, w_r^{(i)}\}_{i=1}^N$ represent N such particles and associated weights for distribution $\pi_r(\boldsymbol{\theta})$ then:

$$\pi_r(\boldsymbol{\theta}) \approx \sum_{i=1}^{N} W_r^{(i)} \delta_{\boldsymbol{\theta}_r^{(i)}}(\boldsymbol{\theta}) \tag{23}$$

where $W_r^{(i)} = w_r^{(i)} \Big/ \sum_{i=1}^{N} w_r^{(i)}$ are the normalized weights and $\delta_{\boldsymbol{\theta}_r^{(i)}}(\cdot)$ is the Dirac function centered at $\boldsymbol{\theta}_r^{(i)}$. Furthermore, for any function $h(\boldsymbol{\theta})$ which is π_r-integrable [7,9]:

$$\sum_{i=1}^{N} W_r^{(i)} h(\boldsymbol{\theta}_r^{(i)}) \to \int h(\boldsymbol{\theta}) \pi_r(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad \text{almost surely} \tag{24}$$

Before discussing the SMC sampler proposed, it is worth recapitulating the basic desiderata:

(a) *Accuracy*: Monte Carlo scheme should be able to correctly sample from *multi-modal* distributions.
(b) *Hierarchical, multiscale*: Monte Carlo scheme should be able to exploit inferences made using forward solvers corresponding to coarser resolutions and refine them as more elaborate forward solvers are used.
(c) *Efficiency*: Monte Carlo sampler should require the fewest possible calls to the forward solver. It should be directly parallelizable and utilize inferences made using cheaper forward solvers corresponding to coarser resolutions in order to reduce the number of calls to more expensive forward solvers corresponding to finer resolutions.

The goal is to obtain samples from each of the posterior distributions in Eq. (21) corresponding to solvers with increasingly finer spatial resolution of the governing PDEs, $r = r_1, r_2, \ldots, r_M$ where r_1 is the coarsest to r_M the finest. For economy of notation we define the artificial posterior $\pi_{r_0}(\boldsymbol{\theta}) = p(\boldsymbol{\theta})$ that coincides with the prior (which is common to all resolutions and independent of the forward solver). To demonstrate the proposed process it suffices to consider a pair of these posterior densities $\pi_1(\boldsymbol{\theta}) \propto L_1(\boldsymbol{\theta}) p(\boldsymbol{\theta})$ and $\pi_2(\boldsymbol{\theta}) \propto L_2(\boldsymbol{\theta}) p(\boldsymbol{\theta})$ corresponding to forward solvers at two successive resolutions r_{i_1} and r_{i_2} (Fig. 2) and discuss the inferential transitions. Let $\pi_{12,\gamma}(\boldsymbol{\theta})$ denote a sequence of artificial, auxiliary distributions defined as follows:

$$\pi_{12,\gamma}(\boldsymbol{\theta}) = \pi_1^{(1-\gamma)}(\boldsymbol{\theta}) \pi_2^{\gamma}(\boldsymbol{\theta}) = L_1^{(1-\gamma)}(\boldsymbol{\theta}) L_2^{\gamma}(\boldsymbol{\theta}) p(\boldsymbol{\theta}) \qquad \gamma \in [0,1] \tag{25}$$

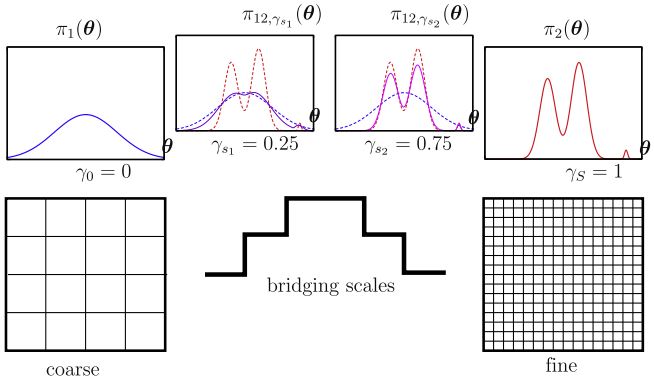


Fig. 2. Illustration of bridging densities as defined in Eq. (25) between posterior distributions $\pi_1(\boldsymbol{\theta}), \pi_2(\boldsymbol{\theta})$ corresponding to different resolutions of the governing PDEs. These allow for accurate and computationally efficient transmission of the inferences made to finer scales.

where γ plays the role of *reciprocal temperature*. Trivially for $\gamma = 0$ we recover π_1 and for $\gamma = 1, \pi_2$. The role of these auxiliary distributions is to *bridge the gap between* π_1 *and* π_2 and provide a smooth transition path where importance sampling can be efficiently applied. In this process, inferences from the coarser scale solver are *transferred and updated* to conform with the finer scale solver. Starting with a particulate approximation for $\pi_{r_0}(\boldsymbol{\theta}) = p(\boldsymbol{\theta})$ (which trivially involves drawing samples from the prior with weights $w_0^{(i)} = 1$), the goal is to gradually update the importance weights and particle locations in order to approximate the target posteriors at various resolutions. In order to implement computationally such a transition we define an increasing sequence of $\{\gamma_s\}_{s=1}^{S}$ with $\gamma_0 = 0$ and $\gamma_S = 1$ (see sub-section 2.2.2). An SMC-based inference scheme would then proceed as described in Table 1.

2.2.1. Notes

- The role of the *Reweighing* step is to correct for the discrepancy between the two successive distributions in exactly the same manner that importance sampling is employed. The *Resampling* step aims at reducing the variance of the particulate approximation by eliminating particles with small weights and multiplying the ones with larger weights. The metric that we use in carrying out this task is the Effective Sample Size (ESS, Table 1) which provides a measure of degeneracy in the population of particles as quantified by their variance. If this degeneracy exceeds a specified threshold, resampling is performed. As it has been pointed out in several studies [14], frequent resampling can deplete the population of its informational content and result in particulate approximations that consist of even a single particle. Throughout this work $ESS_{min} = N/2$ was used. Although other options are available, *multinomial* resampling is most often applied and was found sufficient in the problems examined.
- A critical component involves the perturbation of the population of samples by a standard MCMC kernel in the *Rejuvenation* step as this determines how fast the transition takes place. Although there is freedom in selecting the transition kernel $P_s(\cdot,\cdot)$ (the only requirement is that it is π_{12,γ_s}-invariant), there is a distinguishing feature that will be elaborated further in the next sub-section (see 2.2.3). The target posteriors π_r (as well as the intermediate bridging distributions in Eq. (25)) live in spaces of varying dimensions as previously discussed. Hence an exploration of the state space must involve *trans-dimensional*proposals. Pairs of such moves can be defined in the context of Reversible-Jump MCMC (RJMCMC, [25] such as *adding/deleting* a kernel in the expansion of Eq. (10), or *splitting/merging* kernels (see 2.2.3). Even though it is straightforward to satisfy the invariance constraint in the RJMCMC framework, it is more difficult to design moves that also mix fast. As each (RJ)MCMC requires a likelihood evaluation and a call to a potentially expensive forward solver, it is desirable to minimize their number while retaining good convergence properties.
- In most implementations of such SMC schemes, the sequence of intermediate, bridging distributions is fixed a priori. In order to ensure a smooth transition, a large number is selected at very closely spaced γ_s. It is easily understood that for reasons of computational efficiency, it is desirable to minimize the number of intermediate bridging distributions while ensuring that the successive distributions are not significantly different. In sub-section 2.2.2 we discuss a novel adaptive scheme that allow the automatic determination of these distributions resulting in significant computational savings.
- It should be noted that the framework proposed is directly *parallelizable*, as the evolution (reweighing, rejuvenation) of each particle is *independent* of the rest. Hence the computational effort can be readily distributed to several processors.
- The particulate approximations obtained at each step, provide a *concise* summary of the posterior distribution based on the respective forward solver. This can be readily updated in the manner explained above, if forward solvers at finer resolutions become available or computationally feasible. Similar bridging distributions can be established between distinct forward solvers with differences going beyond their respective resolutions. This is made possible by the *non-parametric Bayesian model* which is independent of the forward solver and the *flexible inference engine based on SMC*.
- An advantageous feature of the proposed framework is that the confidence in the estimates made can be readily quantified by establishing posterior (or credible) intervals, i.e. the posterior probability that the unknown field of interest $f(x)$ exceeds or not a specified threshold, from the particulate approximations (Eq. (23)). It is these credible intervals (or in general measures of the variability in the estimates such as the posterior variance) that can guide *adaptive refinement* of the governing PDEs. Traditionally, adaptive refinement has been based on estimates of some error norm in the

Table 1
Basic steps of an SMC algorithm.

SMC algorithm:

(1) For $s = 0$, let $\{\boldsymbol{\theta}_0^{(i)}, w_0^{(i)}\}_{i=1}^{N}$ be the initial particulate approximation to $\pi_{12,\gamma_0} = \pi_1$. Set $s = 1$.

(2) *Reweigh*: Update weights $w_s^{(i)} = w_{s-1}^{(i)} \frac{\pi_{12,\gamma_s}(\boldsymbol{\theta}_{s-1}^{(i)})}{\pi_{12,\gamma_{s-1}}(\boldsymbol{\theta}_{s-1}^{(i)})}$

(3) *Rejuvenate*: Use an MCMC kernel $P_s(.,.)$ that leaves π_{12,γ_s} invariant to perturb each particle $\boldsymbol{\theta}_{s-1}^{(i)} \rightarrow \boldsymbol{\theta}_s^{(i)}$

(4) *Resample*: Evaluate the Effective Sample Size, $ESS = 1/\sum_{i=1}^{N}(W_{s+1}^{(i)})^2$ and resample the population if it is less than a prescribed threshold ESS_{min}.

(5) The current population $\{\boldsymbol{\theta}_s^{(i)}, w_s^{(i)}\}_{i=1}^{N}$ provides a particulate approximation of π_{12,γ_s} in the sense of Eqs. (23), (24).

(6) If $s < S$ (and $\gamma_s < 1$) then set $s = s + 1$ and goto to step 2. Otherwise stop.

solution of the governing PDEs [1]. It is envisioned that the posterior variance at each point $\boldsymbol{x} \in \mathcal{D}$ in the domain interest can serve as the basis for increasing the resolution of the solver at select regions and making optimal use of the computational resources available.

2.2.2. Bridging distributions π_{12,γ_s}

The role of these auxiliary distributions is to facilitate the transition between two different posteriors π_1 and π_2 corresponding to two distinct solvers. It is easily understood that if π_1 and π_2 are not significantly different, then fewer bridging distributions will be needed and vice versa. As it is impossible to know a priori how pronounced these differences are, in most implementations a rather large number of bridging distributions is adopted, erring on the side of safety.

We propose an adaptive SMC algorithm, that extends existing versions [10,11] in that it automatically determines the number of intermediate bridging distributions needed. In this process we are guided by the Effective Sample Size (ESS, Table 1) which provides a measure of degeneracy in the population of particles. If ESS_s is the *ESS* of the population after the step s and in the most favorable scenario that the next bridging distribution $\pi_{12,\gamma_{s+1}}$ is very similar to π_{12,γ_s}, ESS_{s+1} should not be that much different from ESS_s. On the other hand if that difference is pronounced then ESS_{s+1} could drop dramatically. Hence in order to determine the next auxiliary distribution, we define an acceptable reduction in the *ESS*, i.e. $ESS_{s+1} \geqslant \zeta ESS_s$ (where $\zeta < 1$) and prescribe γ_{s+1} (Eq. (25)) accordingly. The revised Adaptive SMC algorithm is summarized in Table 2.

2.2.3. Trans-dimensional MCMC

As mentioned earlier, a critical component in the SMC framework proposed is the MCMC-based rejuvenation step of the particles $\boldsymbol{\theta}$. It should be noted that the kernel $P_s(\cdot,\cdot)$ in the rejuvenation step (Step 3 of the SMC algorithm) need not be known explicitly as it does not enter in any of the pertinent equations. It is suffices that it is π_{12,γ_s}-invariant which is the target density. For the efficient exploration of the state space, we employ a mixture of moves which involve fixed dimension proposals (i.e. proposals for which the cardinality of the representation k is unchanged) as well as moves which alter the dimension k of the vector of parameters $\boldsymbol{\theta}$. We consider a total of $M = 7$ such moves, each selected with a certain probability as discussed below. Of those, four involve trans-dimensional proposals which warrant a more detailed discussion.

It is generally difficult to design proposals that alter the dimension significantly while ensuring a reasonable acceptance ratio. For that purpose, in this work we consider proposals that alter the cardinality k of the expansion by 1 i.e. $k' = k - 1$ or $k' = k + 1$. We adopt the Reversible-Jump MCMC (RJMCMC) framework introduced in [25] according to which such moves are defined in pairs in order to ensure reversibility of the Markov kernel (even though the reversibility condition is not necessary, it greatly facilitates the formulations). We consider two such pairs of moves, namely *birth-death* and *split-merge*. Let a proposal from $(k, \boldsymbol{\theta})$ to $(k', \boldsymbol{\theta}')$ that increases the dimension i.e. $k' = k + 1$ and $\boldsymbol{\theta} \in \boldsymbol{\Theta}_k, \boldsymbol{\theta}' \in \boldsymbol{\Theta}^{k+1}$ (see last paragraph of sub-Section 2.1.2). Let $p(k \to k')$ the probability that such a proposal is made (user specified) and $p(k' \to k)$ the probability that the *reverse*, dimension-decreasing proposal is made. In order to account for the $m = dim(\boldsymbol{\Theta}_{k+1}) - dim(\boldsymbol{\Theta}_k)$ difference in the dimensions of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, the former is augmented with a vector $\boldsymbol{u} \in \mathbb{R}^m$ drawn from a distribution $q(\boldsymbol{u})$. Consider a differential and one-to-one mapping $h : \boldsymbol{\Theta}_{k+1} \to \boldsymbol{\Theta}_{k+1}$ that connects the three vectors as $\boldsymbol{\theta}' = h(\boldsymbol{\theta}, \boldsymbol{u})$. Then as it is shown in [25], the acceptance ratio of such a proposal is:

$$\min\left\{1, \frac{\pi_{12,\gamma_s}(\boldsymbol{\theta}')p(k \to k')}{\pi_{12,\gamma_s}(\boldsymbol{\theta})p(k' \to k)} \frac{1}{q(\boldsymbol{u})} \left|\frac{\partial \boldsymbol{\theta}'}{\partial(\boldsymbol{\theta}, \boldsymbol{u})}\right|\right\} \tag{26}$$

where $\left|\frac{\partial \boldsymbol{\theta}'}{\partial(\boldsymbol{\theta}, \boldsymbol{u})}\right|$ is the Jacobian of the mapping h. Such a proposal is invariant w.r.t. the density π_{12,γ_s}. Similarly one can define, the acceptance ratio of the *reverse*, dimension-decreasing move:

$$\min\left\{1, \frac{\pi_{12,\gamma_s}(\boldsymbol{\theta})p(k' \to k)}{\pi_{12,\gamma_s}(\boldsymbol{\theta}')p(k \to k')} q(\boldsymbol{u}) \left|\frac{\partial \boldsymbol{\theta}'}{\partial(\boldsymbol{\theta}, \boldsymbol{u})}\right|^{-1}\right\} \tag{27}$$

Details for the reversible pairs used in this work are provided in the Appendix.

Table 2
Basic steps of the *Adaptive* SMC algorithm proposed.

Adaptive SMC algorithm:

(1) For $s = 0$, let $\{\boldsymbol{\theta}_0^{(i)}, w_0^{(i)}\}_{i=1}^N$ be the initial particulate approximation to $\pi_{12,\gamma_0} = \pi_1$ and ESS_0 the associated effective sample size. Set $s = 1$.
(2) *Reweigh*: If $w_s^{(i)}(\gamma_s) = w_{s-1}^{(i)} \frac{\pi_{12,\gamma_s}(\boldsymbol{\theta}_{s-1}^{(i)})}{\pi_{12,\gamma_{s-1}}(\boldsymbol{\theta}_{s-1}^{(i)})}$ are the *updated* weights as a function of γ_s then determine γ_s so that the associated $ESS_s = \zeta ESS_{s-1}$ (the value $\zeta = 0.95$ was used in all the examples). Calculate $w_s^{(i)}$ for this γ_s.
(3) *Resample*: If $ESS_s \leqslant ESS_{min}$ then resample.
(4) *Rejuvenate*: Use an MCMC kernel $P_s(.,.)$ that leaves π_{12,γ_s} invariant to perturb each particle $\boldsymbol{\theta}_{s-1}^{(i)} \to \boldsymbol{\theta}_s^{(i)}$
(5) The current population $\{\boldsymbol{\theta}_s^{(i)}, w_s^{(i)}\}_{i=1}^N$ provides a particulate approximation of π_{12,γ_s} in the sense of Eqs. (23), (24).
(6) If $\gamma_s < 1$ then set $s = s + 1$ and goto to step 2. Otherwise stop.

2.3. Prediction

The significance of mathematical models for the computational simulation of physical processes lies in their *predictive ability*. It is these predictions that serve as the basis for engineering decisions in several systems of technological interest. The proposed framework provides a seamless link from experiments/data collection, to model calibration and ultimately prediction. In the presence of significant sources of uncertainty it is important not only to provide predictive estimates but quantify the level of confidence one can assign to the predicted outcome. The inferred posteriors π_r corresponding to various model resolutions can be used to carry out this task. In accordance with the Bayesian mind-set, all unknowns are considered random. If $\hat{\boldsymbol{y}}$ denotes the output to be predicted (under specified input, boundary and initial conditions) then, the *predictive posterior* $p(\hat{\boldsymbol{y}}|\boldsymbol{y})$ based on the available data $\boldsymbol{y}$ can be expressed as [23]:

$$p(\hat{\boldsymbol{y}}|\boldsymbol{y}) = \int p(\hat{\boldsymbol{y}},\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta} = \int p_r(\hat{\boldsymbol{y}}|\boldsymbol{\theta},\boldsymbol{y}) \underbrace{p(\boldsymbol{\theta}|\boldsymbol{y})}_{\text{posterior}} d\boldsymbol{\theta} = \int \underbrace{L_r(\hat{\boldsymbol{y}}|\boldsymbol{\theta})}_{\text{likelihood}} \pi_r(\boldsymbol{\theta})d\boldsymbol{\theta} \approx \sum_{i=1}^{N} W_r^{(i)} L_r(\hat{\boldsymbol{y}}|\boldsymbol{\theta}_r^{(i)}) \tag{28}$$

The term $p(\hat{\boldsymbol{y}}|\boldsymbol{\theta})$ is the likelihood of the predicted data determined by the forward solver at resolution r as in Eq. (5). Eq. (28) offers an intuitive interpretation of the predictive process. The predictive posterior distribution is a mixture of the corresponding likelihoods evaluated at all possible states $\boldsymbol{\theta}$ of the system, with weights proportional to the their posterior values. In the context of Monte Carlo simulations, samples of $\hat{\boldsymbol{y}}$ from $p(\hat{\boldsymbol{y}}|\boldsymbol{y})$ can be readily drawn using the particulate approximation of each π_r (Eq. (23)). These samples can subsequently be used to statistics of the predicted output $\hat{\boldsymbol{y}}$ such as moments, probabilities of exceedance which can be extremely useful in engineering practice.

3. Numerical examples

The method proposed is illustrated in problems from nonlinear solid mechanics using artificial data. The governing PDEs are those of small-strain, rate-independent, perfect plasticity with a von-Mises yield criterion and associative flow rule [57]:

$$\begin{aligned}
&\nabla\cdot\boldsymbol{\sigma}(\boldsymbol{x}) = 0 \quad \text{(conservation of linear momentum)}\\
&\boldsymbol{\sigma} = \mathbf{C}(E,\nu):(\boldsymbol{\epsilon}-\boldsymbol{\epsilon}^p) \quad \text{(elastic stress—strain relationships)}\\
&h(\boldsymbol{\sigma}) := \sqrt{\|\boldsymbol{\sigma}\|^2 - \frac{1}{3}(tr[\boldsymbol{\sigma}])^2} - \sqrt{\frac{2}{3}}\sigma_{yield} \quad \text{(yield surface)}\\
&\dot{\boldsymbol{\epsilon}}^p = \lambda\frac{\partial h}{\partial\boldsymbol{\sigma}} \quad \text{(flow rule)}
\end{aligned} \tag{29}$$

where $\boldsymbol{\sigma}$ is the Cauchy stress-tensor, $\boldsymbol{\epsilon} = \frac{1}{2}(\nabla u + u\nabla)$ and $\boldsymbol{\epsilon}^p$ the total and plastic-part of the strain tensor, $\boldsymbol{v} = (v_x, v_y, v_z)$ is the displacement vector, $\mathbf{C}(E,\nu)$ is the elastic moduli which depends on the Young's modulus E (it was assumed that it was known $E = 1000$) and Poisson's ratio ν (it was assumed that it was known $\nu = 0.3$). The field of interest in all the problems examined was the yield stress $\sigma_{yield}(\boldsymbol{x})$ which was assumed to vary spatially. The yield stress determines the boundary of the elastic domain in the material response. A square two-dimensional domain $\mathcal{D} = [0,1]\times[0,1]$ under plane stress conditions was considered and the forward solvers were Finite Element models which discretize the governing PDEs of Eq. (29) for $\boldsymbol{x}\in\mathcal{D}$. In order to construct a sequence of solvers operating at different resolutions, we considered 4 different partitions corresponding to uniform $8\times 8, 16\times 16, 32\times 32$ and 64×64 grids (i.e. with element sizes $\frac{1}{8}\times\frac{1}{8}, \frac{1}{16}\times\frac{1}{16}, \frac{1}{32}\times\frac{1}{32}$ and $\frac{1}{64}\times\frac{1}{64}$, respectively). A critical issue with spatially varying parameters is how this variability is accounted in the discretized representation. In this work, we adopted a simple rule according to which each finite element was assigned a constant yield stress value which was equal to the average of the field $\sigma_{yield}(\boldsymbol{x})$ within the element. This scheme by no means represents a consistent upscaling of the governing PDEs let alone being optimal. It can be easily established that it can introduce significant deviations in the effective response which depends on the full details of the spatially varying field. This poor selection is made however to emphasize the point that inaccurate solvers can be useful and can lead to significant improvements in accuracy and efficiency. Their role is to provide a computationally inexpensive approximation to the fine-scale posterior that can be efficiently updated and refined using a reduced number of runs from more expensive solvers. Naturally, if more sophisticated upscaling schemes are introduced, that is procedures that can better represent the subgrid scales (e.g. [12,16,33,34,39,55], the transitions in the sequence of posterior become smoother and the computational effort is further reduced. This is because relatively inexpensive models could still provide us with as much information about the unknown field $f(x)$ as a more expensive (and perhaps more brute force) solver.

Since $\sigma_{yield}(\boldsymbol{x}) > 0\forall\boldsymbol{x}$, we used our model to infer $log(\sigma(\boldsymbol{x}))$ i.e. in Eq. (10), $f(\boldsymbol{x}) = log(\sigma(\boldsymbol{x}))$. The adaptive SMC scheme (Table 2) with $N = 100$ or $N = 500$ particles was employed in the examples presented with $\zeta = 0.95$ and $ESS_{min} = N/2$.

The following values for the hyperparameters of the prior model were used (Section 2.1.2):

- $k_{max} = 100$ and $s = 0.1$ (Eq. (14)).
- $a_\tau = 1.0$ (Eq. (15)) and $a_\mu = 0.0001$ (Eq. (16)).
- $a_0 = 2.0$ and $b_0 = 1\times 10^{-6}$ (Eq. (18)).
- $a = 2.$ and $b = 1.\times 10^{-6}$ (Eq. (5)).

3.1. Example A

In this example it was assumed that the yield stress varied as follows (Fig. 3):

$$\log \sigma_{yield}(\boldsymbol{x}) = -1 \exp\{-10x^2 - 2(y-1)^2\} - 1 \exp\{-2(x-1)^2 - 10y^2\} \tag{30}$$

The nonlinear governing PDEs (Eq. (29)) were solved using a 64×64 uniform finite element mesh with the following boundary conditions:

- $v_x = v_y = 0$ along $x = 0$.
- $v_x = -v_y = 0.001$ along $x = 1$.

The displacements v_x, v_y at a regular grid consisting of 72 points with coordinates $(0.125i, 0.125j)$, for $i = 1, \ldots, 8$ and $j = 0, \ldots, 8$ were recorded resulting in $n = 144$ data points (as in Fig. 3). The empirical mean (of the absolute values) of these observations μ_A was calculated and the recorded values were contaminated by Gaussian noise of standard deviation $5\%\mu_A$ in

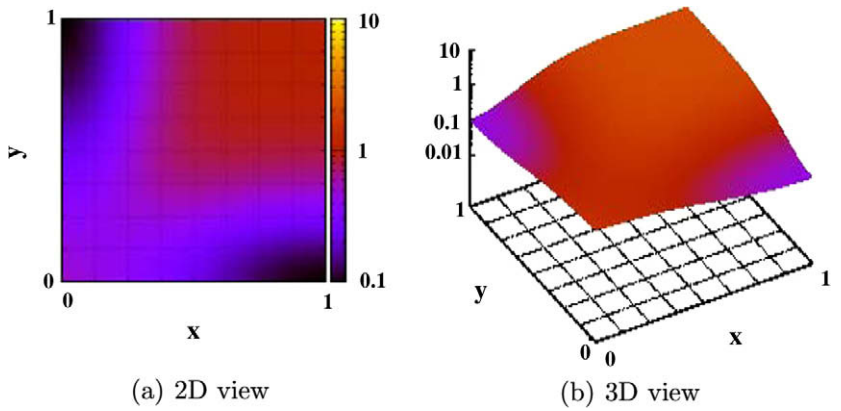


Fig. 3. Reference $\sigma_{yield}(\boldsymbol{x})$ field for Example A.

Table 3
Computational cost of different resolution solvers for Example A.

Solver resolution	Degrees of freedom	Normalized computational time (Actual in s)
16×16	510	$\frac{1}{156}$ (0.55)
32×32	2046	$\frac{1}{18}$ (4.8)
64×64	8190	1 (86)

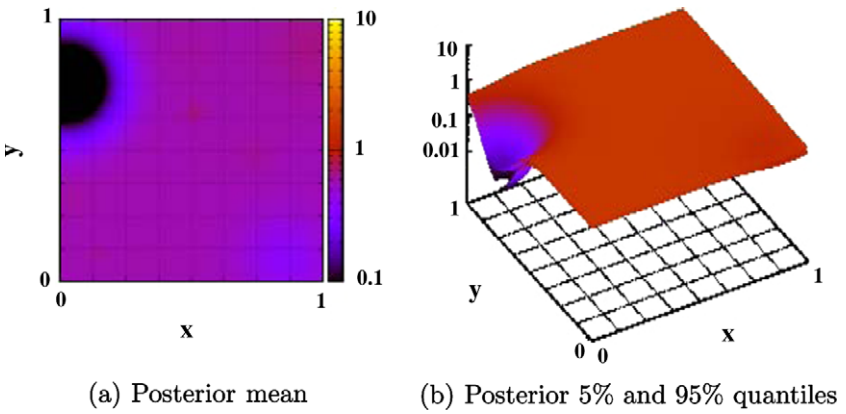
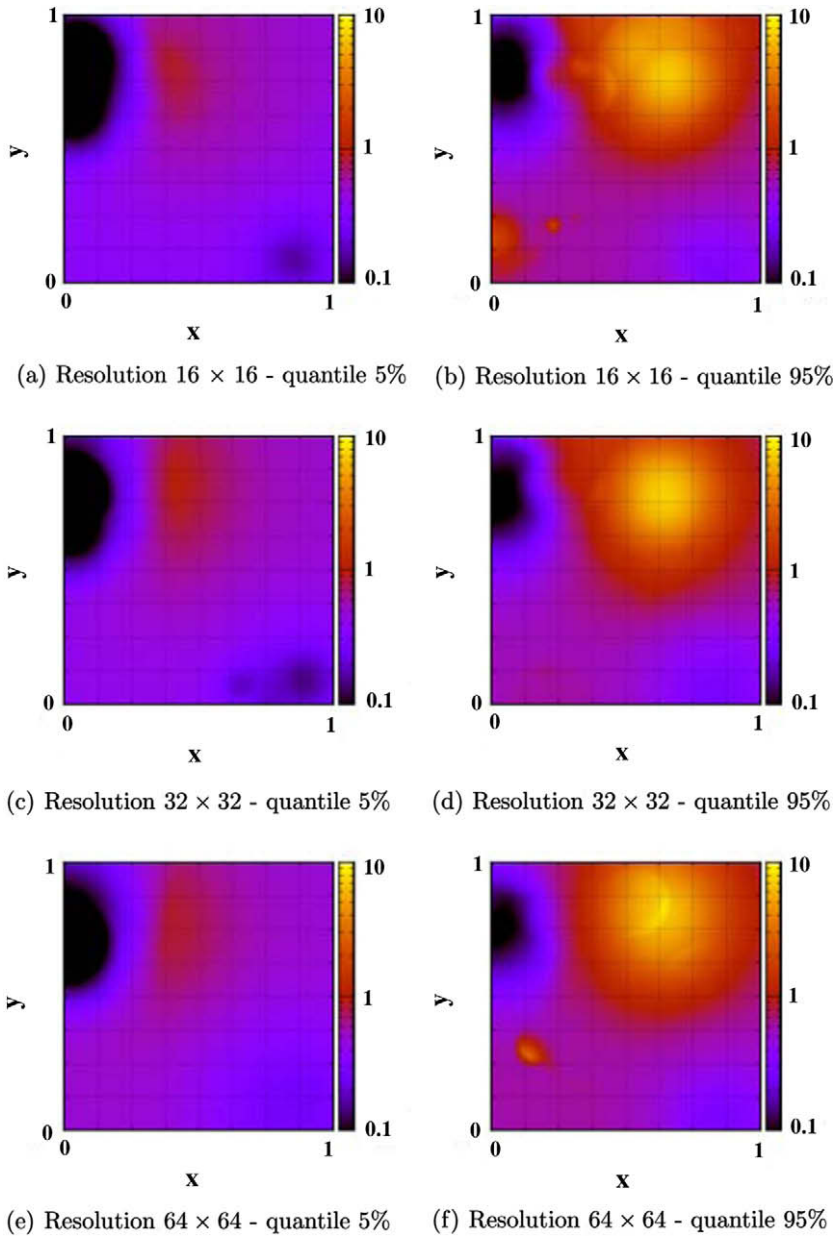


Fig. 4. Posterior inference using only the 64×64 solver.

order to obtain sets of *observables* denoted by $\{y_i\}_{i=1}^n$ in our Bayesian model (Eq. (3)). We note that in this example the scale of variability of the unknown field $\sigma_{yield}(\boldsymbol{x})$ is *larger* than the scale of observations, i.e. the grid size where displacements were recorded.

Table 3 reports the number of degrees of freedom per solver and the normalized computational time for a single run w.r.t. the 64×64 solver. As mentioned earlier, each finite element was assigned a constant yield stress equal to the average value inside the element. This is of course inconsistent with the governing PDEs as the geometry of the variability plays a critical role for the effective properties of each element. It is easily understood though that the corresponding posterior should have some similarities arising from the mere nature of their construction.

At first, we attempted to solve the problem by operating solely on the finest solver. Using the Adaptive SMC scheme proposed with $N = 100$ particles, this resulted in a sequence of 163 (between the prior π_0 and the target posterior) auxiliary bridging distributions constructed as mentioned earlier. The inferred field (posterior mean and quantiles) are depicted in Fig. 4. Even though they exhibit similarities with the ground truth (Fig. 3), there are also considerable differences which suggest that the algorithm probably got trapped in some mode of the posterior. This is to be expected due to the highly



(a) Resolution 16×16 - quantile 5% (b) Resolution 16×16 - quantile 95%

(c) Resolution 32×32 - quantile 5% (d) Resolution 32×32 - quantile 95%

(e) Resolution 64×64 - quantile 5% (f) Resolution 64×64 - quantile 95%

Fig. 5. Posterior quantiles at various solver resolutions for Example A.

nonlinear nature of the forward solver and the large state space. It is possible however that the correct solution could be recovered if the size of the population and/or the number of bridging distributions is increased. Inspite of that, it is the significant computational effort that makes such an approach impractical. In particular 16,300 (i.e. 163×100) calls to the most expensive forward solver were required.

In contrast, when a sequence of three solvers was used the results obtained are significantly closer to the ground truth as it can be seen in Figs. 5 and 6. It is observed that even using the coarsest solver (16×16), we are able to correctly identify some of the basic features of the underlying field. The inferences are greatly improved as solvers at finer resolutions are invoked. Fig. 7 depicts the number of bridging distributions needed at each resolution and the respective reciprocal temper-

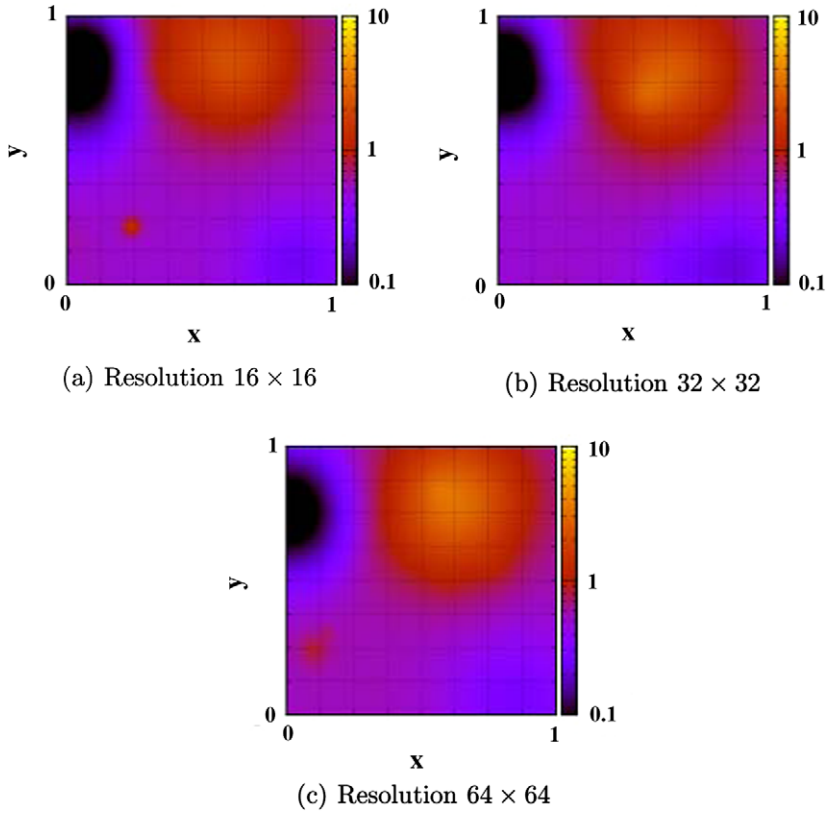


Fig. 6. Posterior mean at various solver resolutions for Example A.

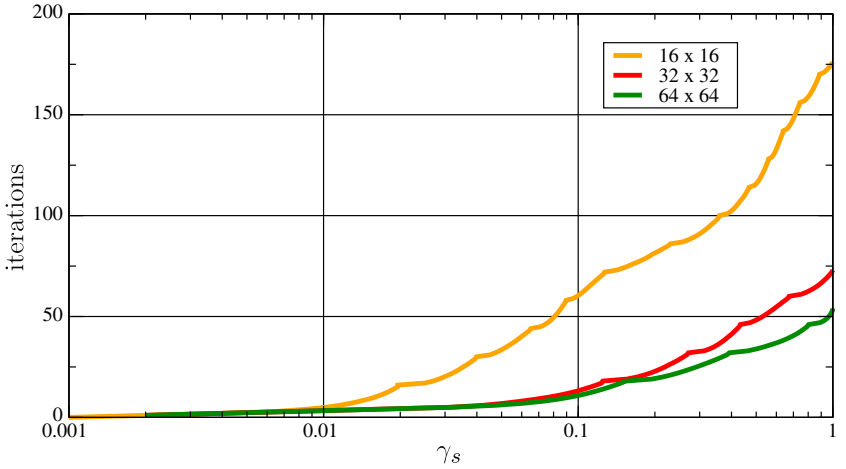
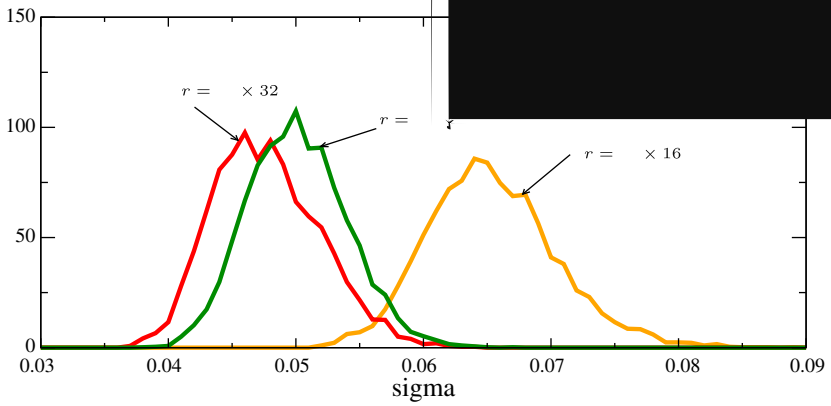
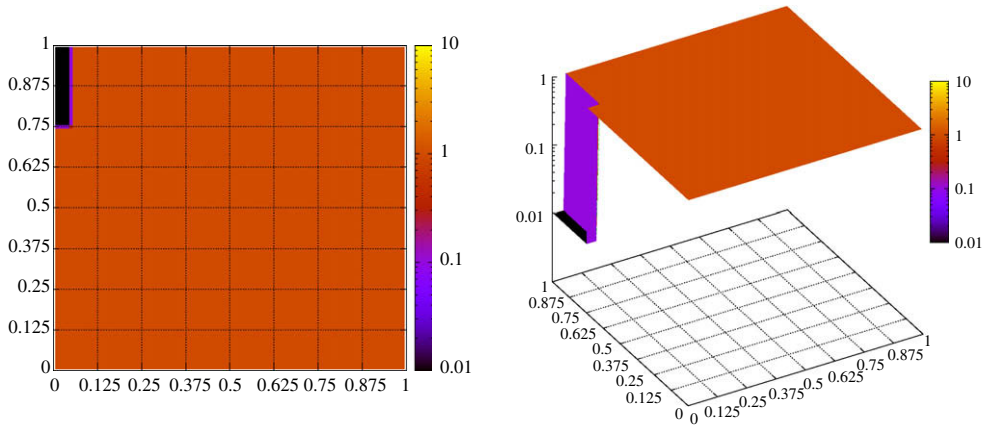


Fig. 7. Evolution of reciprocal temperature γ_s (Eq. (25)) and number of bridging distributions.

atures γ_s (Eq. (25)). These were *automatically determined* by the
served that the number of intermediate distributions needed d
consequence of the ability of the proposed scheme to accumul
summarized in Table 4 which also reports the *effective* comput
that a reduction of the total number of calls is achieved (16,30

Fig. 8 depicts the posterior densities of the inferred error stan
the proposed technique is able to quantify the magnitude of the
for the reference resolution 64×64 it correctly detects that the
depicts the marginal posterior on k, i.e. the cardinality of the ex
method leads to *sparse* representations (on average $k = 5$ and the
the accuracy. Traditional formulations (deterministic or probabil
the 64×64 mesh, 4096 parameters) and therefore require operat
implications this carries.





(a) 2D view (b) 3D view

Fig. 10. Reference $\sigma_{yield}(\boldsymbol{x})$ field for Example B.

3.2. Example B

In this example the yield stress $\sigma_{yield}(\boldsymbol{x})$ was assumed to vary as follows:

$$\sigma_{yield}(\boldsymbol{x}) = \begin{cases} 0.01, & \text{if} \quad \boldsymbol{x} \in [0.0, 0.05] \times [0.75, 1.0] \\ 1.0, & \text{otherwise} \end{cases} \tag{31}$$

As in the previous example, the governing PDEs were solved using a 64×64 uniform finite element mesh with the following boundary conditions for the horizontal v_x and vertical v_y displacements:

- $v_x = v_y = 0$ along $x = 0$.
- $v_x = -v_y = 0.001$ along $x = 1$.

The displacements v_x, v_y at a regular grid consisting of 72 points with coordinates $(0.125i, 0.125j)$, for $i = 1, \ldots, 8$ and $j = 0, \ldots, 8$ were recorded resulting in $n = 144$ data points (Fig. 10).

The empirical mean of (the absolute values) of these observations μ_B was calculated and the recorded values were contaminated by Gaussian noise of standard deviation $5\%\mu_B$ and $10\%\mu_B$ in order to obtain two sets of *observables* denoted by $\{y_i\}_{i=1}^n$ in our Bayesian model (Eq. (3)), i.e.:

- Dataset B1 contaminated with $5\%\mu_B$ noise.
- Dataset B2 contaminated with $10\%\mu_B$ noise.

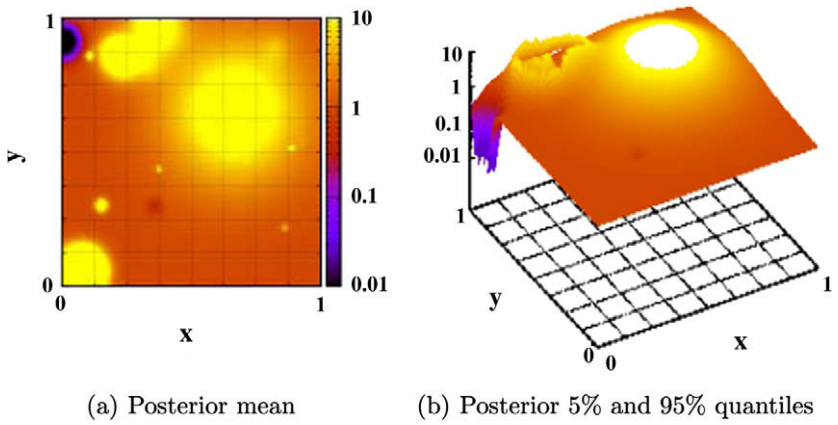
In contrast to the previous problem, the scale of variability of $\sigma_{yield}(\boldsymbol{x})$ is *smaller* than the scale of observations, i.e. the grid size where displacements were recorded. In principle therefore, the data does not sufficiently resolve the material inhomogeneity.

We considered three different forward solvers (and therefore three posteriors π_r) corresponding to a $16 \times 16, 32 \times 32$ and 64×64 uniform Finite Element discretization. Table 5 reports the number of degrees of freedom per solver and the normalized computational time for a single run w.r.t. the 64×64 solver. As mentioned earlier, each finite element was assigned a constant yield stress equal to the average value inside the element.

At first, we attempted to solve the problem with non-contaminated data and operating directly on the finest solver and trying to sample from a single posterior. The Adaptive SMC scheme proposed with $N = 500$ particles resulted in a sequence of 215 auxiliary bridging distributions constructed as mentioned earlier. The algorithm failed to converge to something reasonably close to the reference solution (Fig. 11) as it got trapped in some mode of the posterior. This is to be expected due to

Table 5
Computational cost of different solvers.

Solver resolution	Degrees of Freedom	Normalized computational time (Actual in s)
16×16	510	$\frac{1}{120}$ (1.5)
32×32	2046	$\frac{1}{36}$ (5.0)
64×64	8190	1 (180)

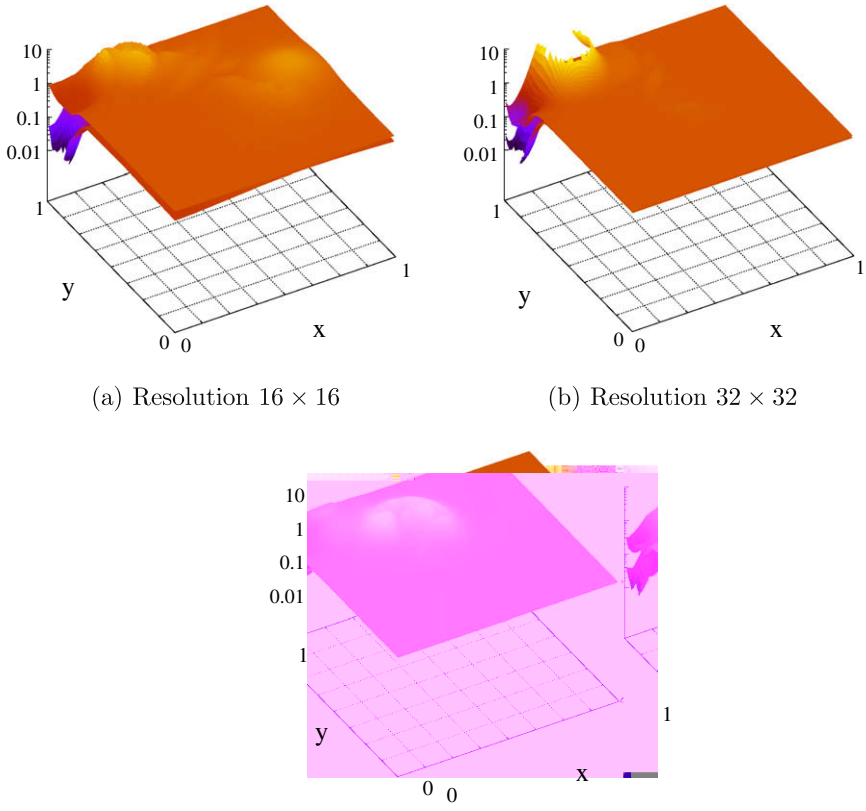


(a) Posterior mean (b) Posterior 5% and 95% quantiles

Fig. 11. Posterior inference using only π_r corresponding to 64×64 solver.

the highly nonlinear nature of the forward solver and the large state space. It is possible however that the correct solution could be recovered if the size of the population and/or the number of bridging distributions is increased. Despite that, the associated computational effort that makes such an approach impractical. In particular 107,500 (i.e. 215×500) calls to the most expensive forward solver were required.

We employed the algorithm proposed using $N = 500$ particles and the sequence of posteriors implied by the increasingly finer solvers and for datasets B1 and B2. Figs. 12 and 13 depict posterior quantiles for B1 and B2, respectively. Despite the contamination of the reference data, the algorithm is in both cases able to identify fields that are significantly closer to the ground truth as compared with the inferences above obtained for data with no error.



(a) Resolution 16×16 (b) Resolution 32×32

(c) Resolution 64×64

Fig. 12. Posterior quantiles 5% and 95% for inferred $\sigma_{yield}(\boldsymbol{x})$ from dataset B1.

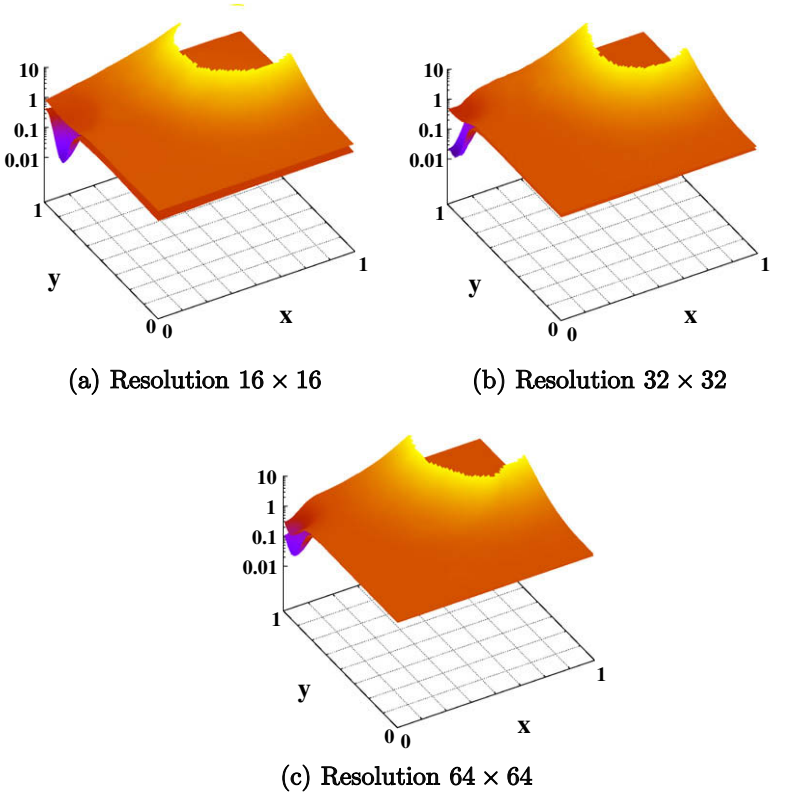


Fig. 13. Posterior quantiles 5% and 95% for inferred $\sigma_{yield}(\boldsymbol{x})$ from dataset B2. Note then even though the 95% quantile is much larger than the reference in the upper-right hand corner, the 5% quantile correctly bounds it from below.

Table 6
Computational cost for inferences based on dataset B1.

Solver resolution	Number of bridging distributions	Computational effort (w.r.t. calls to 64×64 solver)
16×16	122	508
32×32	75	1354
64×64	64	32,889
Total		34,751

Table 7
Computational cost for inferences based on dataset B2.

Solver resolution	Number of bridging distributions	Computational effort (w.r.t. calls to 64×64 solver)
16×16	91	379
32×32	45	813
64×64	23	11,819
Total		13,011

More importantly though these inferences are obtained at a fraction of the computational cost as summarized in Tables 6 and 7. Fig. 14 depicts the posterior densities of the inferred model error standard deviations σ_r described in Eq. (4). It is readily seen that the proposed technique is able to quantify the magnitude of the model error for solvers of various resolutions. Furthermore for the reference resolution 64×64 it correctly detects that the error contamination is of the level of $5\%\mu_B$ and $10\%\mu_B$ for dataset B1 and B2, respectively. Fig. 15 depicts posterior statistics of the inferred $\sigma_{yield}(\boldsymbol{x})$ field along the line $x = 0$. It is noted that the true value is detected for $y < 0.75$ (or at least is contained in the credible intervals obtained) but significant differences are observed for $y \geqslant 0.75$. This might seem at first as an erroneous result and a failure of the proposed scheme. In reality though, we were able to identify configurations that are consistent with the observables and consistent with the level of noise in the data. It should come as no surprise that the available data (even in the absence of noise) are

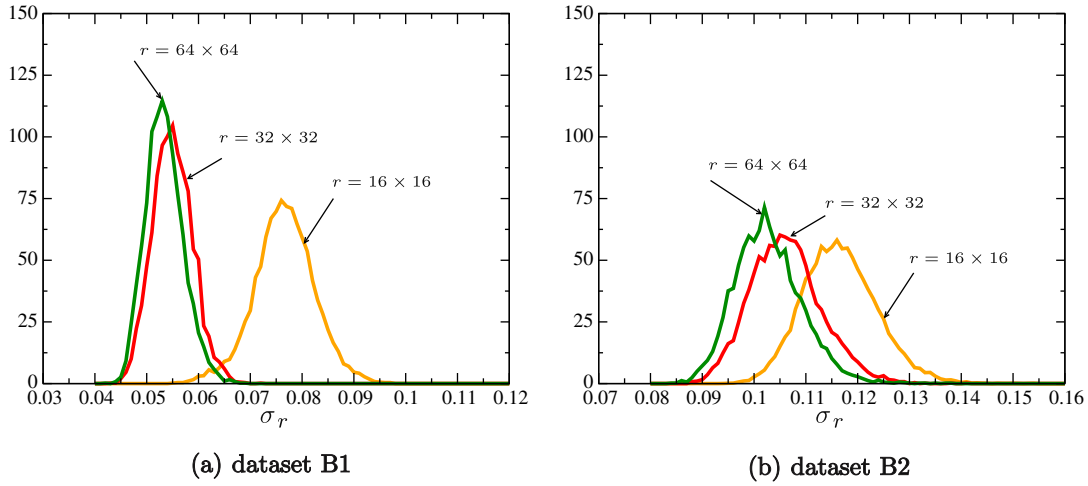


Fig. 14. Posterior densities of model error std. deviations σ_r as in Eq. (4). (The values on x-axis have been divided by μ_B).

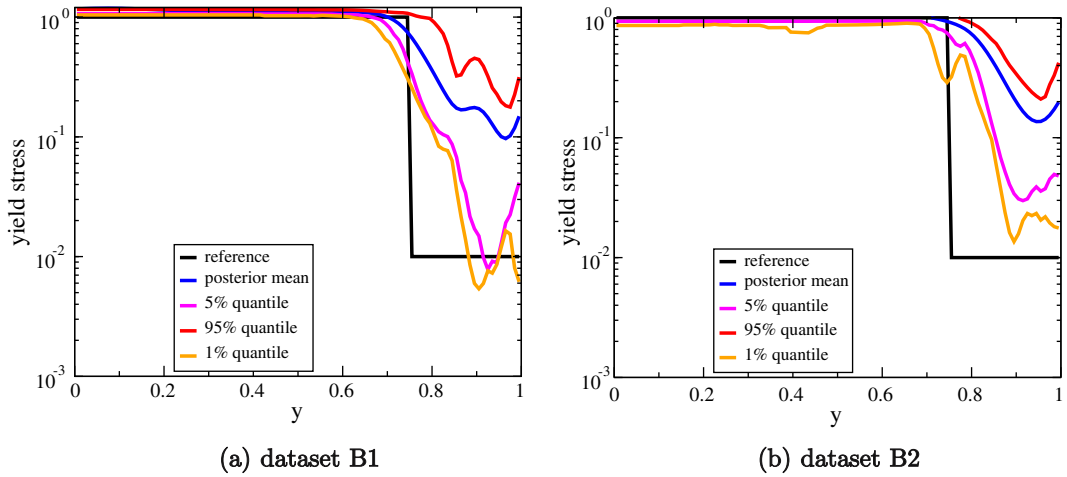


Fig. 15. Posterior statistics of $\sigma_{yield}(x)$ along $x = 0$ for solver at resolution 64×64.

not sufficient to uniquely identify the underlying material properties and especially since the scale of material variability is smaller than the data grid. There are several configurations equally or comparably compatible as the ground truth. Amongst all these configurations, the Bayesian scheme proposed favors the ones that are more *plausible* in the sense of having higher posterior values. Given the parsimonious construction of the prior model, favoring sparser representations and slower-varying fields, the proposed algorithm identified the *simplest* configuration that explains the data. This naturally raises the question what should be a proper validation metric of the proposed and other system identification techniques. Given the ill-posedeness of the problem in the sense explained earlier, we argue herein that validation should be based on the *predictive ability* of the identified model. Hence if the inferred material property field(s) can give rise to response predictions consistent with the predictions of the actual field, then they are equivalent. This is also consistent with the basic functionality of computational forward models and associated parameters as predictive tools of the output/response/performance under various inputs/excitations and initial/boundary conditions. In order to assess the predictive ability of the inferred field and associated model, we provide samples of the predictive posterior for the displacements at the center of the domain under a different loading regime than the one for which the *training data* was obtained. In particular the following boundary conditions were used:

- Displacements $v_x = v_y = 0$ along $x = 0$.
- Normal traction $v_x = 0.002$ along $x = 1.0$.

The samples of the predictive posterior were obtained as discussed in Section 2.3 and are depicted in Fig. 16. It is noted that even for the models corresponding to coarser resolutions, we are able to predict output consistent with the reference

are considered. As one would expect, the associated uncertainty is larger when inferences were based on dataset B2 as the latter is contaminated with higher noise levels. Similar conclusions can be drawn when looking at the predictive posterior (Fig. 17) for v_y at $(x = 0.125, y = 1.0)$ very close to the indentation that was *inaccurately identified* according to Fig. 15. In all cases the reference value is within the range of predicted outcomes. Similar results were obtained for the response at several other locations in the specimen and loading conditions, but are omitted due to space restrictions.

3.3. Example C

The final numerical illustration involved an elasto-plastic material model as well occupying the unit square domain $\mathcal{D} = [0,1] \times [0,1]$. It was assumed that the yield stress $\sigma_{yield}(\boldsymbol{x})$ varied spatially as shown in Fig. 18. The particular field is a realization of a non-Gaussian random field $Z(\boldsymbol{x})$ defined as:

$$Z(\boldsymbol{x}) = 1.01 - \exp\{-Y^2(\boldsymbol{x})\} \tag{32}$$

where $Y(\boldsymbol{x})$ is a homogeneous, Gaussian random field with zero mean, unit variance and isotropic correlation $\rho(r) = \exp\{-r/0.5\}$. By definition any sample realization of $Z(\boldsymbol{x})$ takes values in [0.01, 1.01).

The following boundary conditions were imposed:

- $v_x = v_y = 0$ along $x = 0$.
- $v_x = -v_y = \xi$ 0.001 along $x = 1$.

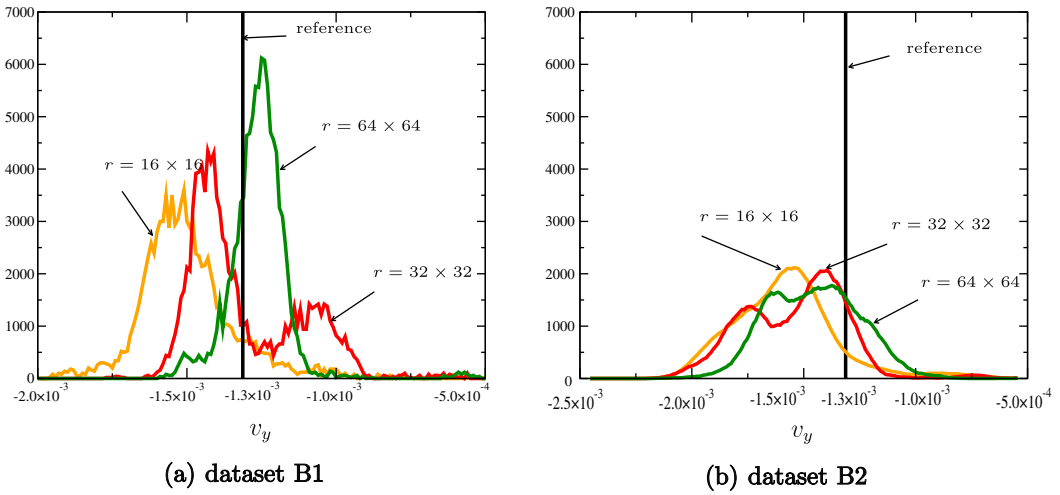


Fig. 17. Predictive posterior for vertical displacement v_y at (0.125, 1.0).

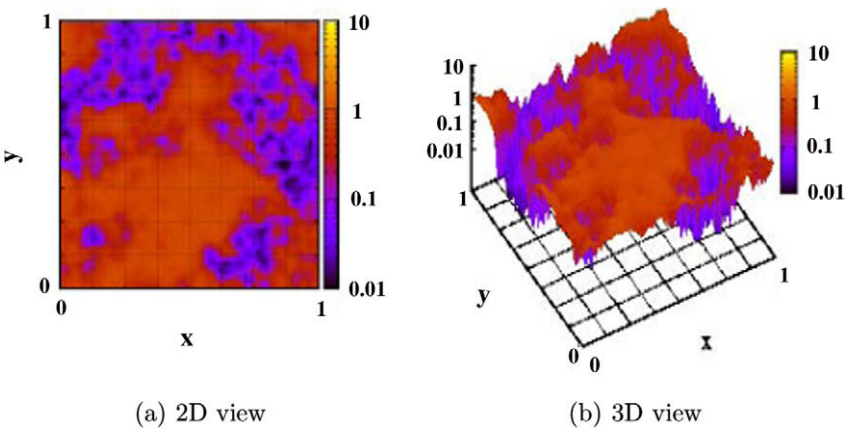


Fig. 18. Reference $\sigma_{yield}(\boldsymbol{x})$ field for Example C.

The quasi-static governing PDEs were solved using a uniform 64×64 finite element mesh for $\xi \in [0,1]$. The displacements v_x, v_y at a regular grid consisting of 72 points with coordinates $(0.125i, 0.125j)$, for $i = 1,\ldots,8$ and $j = 0,\ldots,8$ were recorded at 10 values of $\xi = k0.1, k = 1,2,\ldots,10$ resulting in a total of $n = 144 \times 10 = 1,440$ data points, say $\{\hat{y}_i\}_{i=1}^{n}$. This data were contaminated as follows:

$$y_i = \hat{y}_i(1 + \sigma_y \eta_i) \tag{33}$$

where $\eta_i \sim N(0,1)$ (i.i.d) and $\sigma_y = 0.05$. It was the resulting $\{y_i\}_{i=1}^{n}$ that served as the *observables* in our Bayesian model (Eq. (3)). This is particularly challenging example as the scale of variability of $\sigma_{yield}(\boldsymbol{x})$ is *much smaller* than the scale of observations, i.e. the grid size where displacements were recorded as it can be readily seen in Fig. 18.

We considered four different forward solvers (and therefore four posteriors π_r) corresponding to a $8 \times 8, 16 \times 16, 32 \times 32$ and 64×64 uniform Finite Element discretizations. Table 8 reports the number of degrees of freedom per solver and the normalized computational time for a single run w.r.t. the 64×64 solver. As with previous examples, each finite element was assigned a constant yield stress equal to the average value of $\sigma_{yield}(x)$ inside the element.

Fig. 19 depicts the posterior means for solvers at various resolutions. Table 9 summarizes the associated computational cost. It is noted that even using the coarsest solver (8×8) which amounted to an equivalent of 147 runs to the finest solver (64×64), one can get a good idea of the underlying property (compare Fig. 18 with Fig. 19). These inferences are refined by using the information provided by finer resolution (and more expensive) forward solvers.

In order to assess the predictive ability of the inferred field and associated model, we provide samples of the predictive posterior for the displacements at the center of the domain under a different loading regime than the one for which the *training data* was obtained. In particular the following boundary conditions were used:

Table 8
Computational cost of different solvers.

Solver resolution	Degrees of freedom	Normalized computational time (actual in s)
8×8	126	$\frac{1}{1,588}$ (0.17)
16×16	510	$\frac{1}{321}$ (0.84)
32×32	2046	$\frac{1}{37}$ (7.3)
64×64	8190	1 (270)

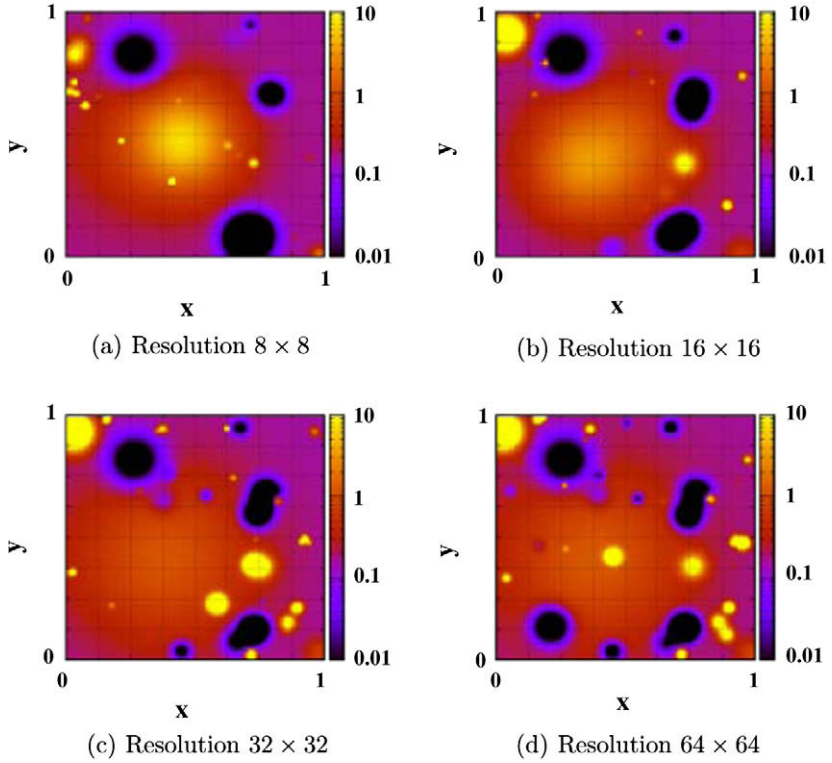
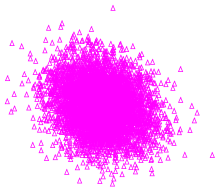


Fig. 19. Posterior mean for various resolutions.

- Displacements $v_x = v_y = 0$ along $x = 0$.
- Displacement $v_x = 0.002$ along $x = 1.0$.

The samples of the predictive posterior were obtained as discussed in Section 2.3 and are depicted in Fig. 20. In all cases the reference value is within the range of predicted outcomes. Furthermore, the proposed scheme provides quantitative measures of the *predictive uncertainty*. Finally Fig. 21 shows the marginal posterior on the size of the field expansion k. It is noted that as finer resolution solvers are employed, the cardinality increases in order to account for the additional configurational details that these solvers provide.



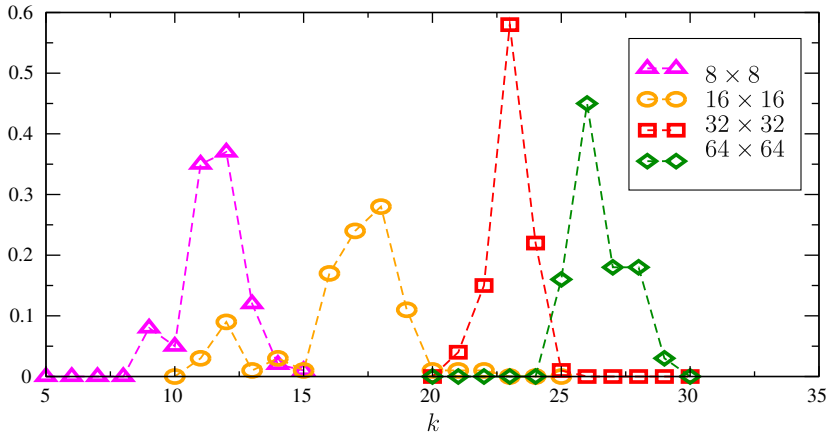


Fig. 21. Posterior for the cardinality k of the field representation.

4. Conclusions

A general Bayesian framework has been presented for the identification of spatially varying model parameters. The proposed model utilizes a parsimonious, non-parametric formulation that favors sparse representations and whose complexity can be determined from the data. An efficient inference scheme based on SMC has been discussed which is embarrassingly parallelizable and well-suited for detecting multi-modal posterior distributions. They key element is the introduction of an appropriate sequence of posteriors based on a natural hierarchy introduced by various forward solver resolutions. As a result, inexpensive, coarse solvers are used to identify the most salient features of the unknown field(s) which are subsequently enriched by invoking solvers operating at finer resolutions. The overall computational cost is further reduced by employing a novel adaptive scheme that automatically determines the number of intermediate steps. The proposed methodology does not require that Markov Chains using all the solvers to be run simultaneously as in other multi-resolution formulations [30] . The particulate approximations provide a concise way of representing the posterior which can be readily updated if the analyst wants to employ forward models operating at even finer resolutions or in general more accurate solvers. The output of the inference algorithm provides estimates of the model error or noise contained in the data. An important feature is the ability to readily provide not only predictive estimates but also quantitative measures of the *predictive uncertainty*. Hence it offers a seamless link between data, computational models and predictions. The efficiency of the sampling schemes proposed could be greatly improved if the proposed moves incorporate information about the governing PDEs and if upscaling relations are available. A feature that was not explored in the examples presented is the possibility of performing *adaptive refinement*, not for the purposes of improving the forward solver accuracy but rather for increasing the resolution of the unknown fields. This can be achieved in two ways and is a direct consequence of the ability of the proposed model (and Bayesian models in general) to produce credible intervals for the estimates made at each step. Hence in regions where the variance of the estimates (or some other measure of random variability) is high, the resolution of the forward solver can be increased. Furthermore, additional measurements/data can be obtained at these regions if such a possibility exists. Hence the proposed framework allows for near-optimal use of the computational resources and sensors available.

Acknowledgment

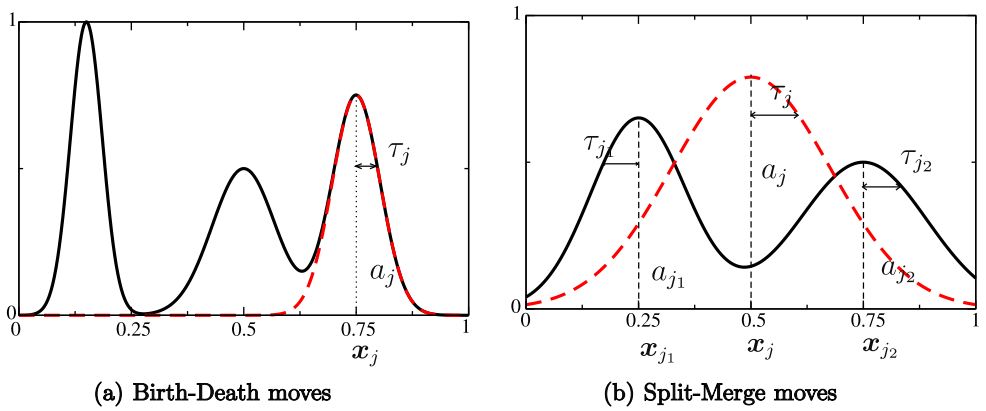
The author would like to acknowledge the two anonymous reviewers whose valuable comments and suggestions improved the overall quality of the paper.

Appendix A

The Appendix provides details relating to the pairs of reversible updates used in the context of RJMCMC (Section 2.2).

Birth–Death

In order to simplify the resulting expressions, we assign the following probabilities of proposing one of these moves $p_{birth} = cmin\left\{1, \frac{p(k+1)}{p(k)}\right\} = c\frac{1}{s+1}$ (from Eq. (14)) and $p_{death} = cmin\left\{1, \frac{p(k-1)}{p(k)}\right\} = c$ (from Eq. (14)). The constant c is user-specified (it is taken equal to 0.2 in this work). Obviously if $k = k_{max}, p_{birth} = 0$ and if $k = 0, p_{death} = 0$ (Fig. 22a).



(a) Birth-Death moves (b) Split-Merge moves

Fig. 22. Trans-dimensional RJMCMC proposals.

For the death move:

- A kernel $j (1 \leqslant j \leqslant k)$ is selected uniformly and removed from the representation in Eq. (10).
- The corresponding a_j is also removed.

For the birth move:

- A new kernel $k+1$ is added to the expansion while the existing terms remain unaltered.
- The associated amplitude a_{k+1} is drawn from $\mathcal{N}(0, \sigma_4^2)$ (the variance σ_4^2 is equal to the average of the squared amplitudes a_j over all the particles at the previous iteration).
- The associated scale parameter τ_{k+1} is drawn from the prior, Eq. (16).
- The associated kernel location $\boldsymbol{x}_{k+1}$ is also drawn from the prior, Eq. (19).

Hence the vector of dimension-matching parameters $\boldsymbol{u}$ consists of $\boldsymbol{u} = (a_{k+1}, \tau_{k+1}, \boldsymbol{x}_{k+1})$ and the corresponding proposal $q(\boldsymbol{u})$ is:

$$q(\boldsymbol{u}) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_4} e^{-\frac{1}{2}a_{k+1}^2/\sigma_4^2} \frac{b_\tau^{a_\tau}}{\Gamma(a_\tau)} \tau_{k+1}^{a_\tau - 1} \exp(-b_\tau \tau_{k+1}) \frac{1}{|\mathcal{D}|} \tag{34}$$

It is obvious that the Jacobian of such a transformation is 1.

Split-merge

These moves correspond to splitting an existing kernel into two or merging two existing kernels into one (Fig. 22b). Similarly to the birth-death pair, they alter the dimension of the expansion by 1 and are selected with probabilities $p_{split} = \frac{1}{s+1}$ and $p_{merge} = c$. For obvious reasons, $p_{split} = 0$ if $k = k_{max}$ and $p_{merge} = 0$ if $k \leqslant 1$. Consider first the merge move between two kernels j_1 and j_2. In order to ensure a reasonable acceptance ratio, merge moves are only permitted when the (normalized) distance between the kernels is relatively small and when the amplitudes a_{j_1}, a_{j_2} are relatively similar. Specifically we require that the following two conditions are met:

$$\frac{\|\boldsymbol{x}_{j_1} - \boldsymbol{x}_{j_2}\|}{\sqrt{\tau_{j_1}^{-1} + \tau_{j_2}^{-1}}} \leqslant \delta_x \quad |a_{j_1} - a_{j_2}| \leqslant \delta_a \tag{35}$$

(the values $\delta_x = \delta_a = 1$ were used in this work). Two candidate kernels are selected uniformly from the pool of pairs satisfying the aforementioned conditions. The proposed kernels j_1 and j_2 are removed from the expansion and are substituted by a new kernel j with the following associated parameters:

- $\tau_j = (\tau_{j_1}^{-1} + \tau_{j_2}^{-1})^{-1}$ (36)
- $a_j = \sqrt{\tau_j}\left(\frac{a_{j_1}}{\sqrt{\tau_{j_1}}} + \frac{a_{j_2}}{\sqrt{\tau_{j_2}}}\right)$ (37)

This ensures that the *average* value of the previous expansion (with j_1 and j_2) in Eq. (10) when integrated in $\mathbb{R}^d$ is the same with the new (which contains j in place of j_1 and j_2)

- $\boldsymbol{x}_j = \dfrac{\boldsymbol{x}_{j_1} + \boldsymbol{x}_{j_2}}{2}$ (38)

The split move is applied to a kernel j (selected *uniformly*) which is substituted by two new kernels j_1, j_2. In order to ensure *reversibility*, kernels j_1 and j_2 should satisfy the requirements of Eq. (35) and the application of a merge move in the manner described above, should return to the original kernel j. There are several ways to achieve this, corresponding essentially to different vectors $\boldsymbol{u}$ and mappings h in Eq. (26). In this work:

- A scalar u_τ is drawn from the uniform distribution $U[0,1]$ and $\tau_{j_1}^{-1} = u_\tau \tau_j^{-1}$ and $\tau_{j_2}^{-1} = (1-u_\tau)\tau_j^{-1}$. This ensures compatibility with Eq. (36).
- A vector $\boldsymbol{u}_x$ is drawn uniformly in the ball of radius R where $R = \frac{\delta_x}{2\sqrt{\tau_j}}$. The center of the new kernels are specified as $\boldsymbol{x}_{j_1} = \boldsymbol{x}_j - \boldsymbol{u}_x$ and $\boldsymbol{x}_{j_2} = \boldsymbol{x}_j + \boldsymbol{u}_x$. This ensures compatibility with the first of Eq. (35) as well as Eq. (38).
- A scalar u_a is drawn from the uniform distribution $U\left[-\frac{\delta_a}{2}, \frac{\delta_a}{2}\right]$. The amplitudes of the new kernels are determined by $a_{j_1} = \hat{a} - u_a$ and $a_{j_2} = \hat{a} + u_a$, where $\hat{a} = \frac{a + u_a\left(\sqrt{u_\tau} - \sqrt{1-u_\tau}\right)}{\sqrt{u_\tau} + \sqrt{1-u_\tau}}$. This ensures compatibility with the second of Eq. (35) as well as Eq. (37).

The vector of dimension-matching parameters $\boldsymbol{u}$ (in Eq. (26)) consists of $\boldsymbol{u} = (u_\tau, \boldsymbol{u}_x, u_a)$ and the corresponding proposal $q(\boldsymbol{u})$ is a product of uniforms in the domains specified above. After some algebra, it can be shown that the Jacobian of such a transformation is $2^{d+1} \frac{\tau}{u_\tau^2 (1-u_\tau)^2} \frac{1}{\sqrt{u_\tau} + \sqrt{1-u_\tau}}$.

The remaining three proposals, involve fixed-dimension moves that do not change the cardinality of the expansion but rather perturb some of the terms involved. In particular, we considered updates of the amplitude a_j, scale τ_j or location $\boldsymbol{x}_j$ of a kernel j selected *uniformly* (naturally, in the case of the amplitudes, the constant a_0 (Eq. (10)) is also a candidate for updating). Each of these three moves is proposed with probability $\frac{1}{3}(p_{birth} + p_{death} + p_{split} + p_{merge}) = \frac{2c}{3}\left(\frac{1}{s+1} + 1\right)$. In particular:

(1) Update $a_j \to a_j'$: A coefficient a_j (in Eq. (10)) is *uniformly* selected and perturbed as:

$$a_j' = a_j + \sigma_1 Z, \quad Z \sim \mathcal{N}(0,1) \tag{39}$$

(2) Update $\tau_j \to \tau_j'$: A scale parameter τ_j (in Eq. (10)) is *uniformly* selected and perturbed as:

$$\tau_j' = \tau_j e^{\sigma_2 Z}, \quad Z \sim \mathcal{N}(0,1) \tag{40}$$

(this ensures positivity of τ_j')

(3) Update $\boldsymbol{x}_j \to \boldsymbol{x}_j'$: A location $\boldsymbol{x}_j \in \mathcal{D} \subset \mathbb{R}^d$ (in Eq. (10)) is *uniformly* selected and perturbed as:

$$\boldsymbol{x}_j' = \boldsymbol{x}_j + \sigma_3 \boldsymbol{Z}, \quad \boldsymbol{Z} = (Z_1, \ldots, Z_d), \quad Z_i \sim \mathcal{N}(0,1) \tag{41}$$

The acceptance ratios are calculated based on the standard MCMC formulas using π_{12,γ_s} as the target density. It should be noted that the variances in the random walk proposals are adaptively selected so that the respective acceptance rates are in the range $0.2 - 0.4$. As it is well-known (chapter 7.6.3 in [54] adaptive adjustments of Markov Chains based on past samples can breakdown ergodic properties and lead to convergence issues in standard MCMC contexts. In the proposed SMC framework however, such restrictions do not apply as it suffices that the MCMC kernel is invariant. This is an additional advantage of the proposed simulation scheme in comparison to traditional MCMC.

References

[1] M. Ainsworth, J.T. Oden, A Posterori Error Estimation in Finite Element Analysis, Wiley-Interscience, 2000.
[2] K.E. Andersen, S.P. Brooks, M.B. Hansen, Bayesian inversion of geoelectrical resistivity data, Technical Report, Dept. Math. Sci., Aalborg University, 2001.
[3] R.P. Barry, J.M. Ver Hoef, Blackbox kriging: spatial prediction without specifying variogram models, Journal of Agricultural Biological and Environmental Statistics 1 (1996) 297–322.
[4] J.C. Brigham, W. Aquino, F. Mitri, J.F. Greenleaf, M. Fatemi, Inverse estimation of viscoelastic material models for solids immersed in fluids using acoustic emissions, Journal of Applied Physics 101 (2007).
[5] D. Calvetti, L. Reichel, Tikhonov regularization of large linear problems, BIT Numerical Mathematics 43 (2) 2003 263–283.
[6] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Review 43 (2001) 129–159.
[7] N. Chopin, Central limit theorem for sequential monte carlo methods and its application to bayesian inference, Annals of Statistics 32 (6) (2004) 2385–2411.
[8] P.S. Craig, M. Goldstein, J.C. Rougier, A.H. Seheult, Bayesian forecasting for complex systems using computer simulators, Journal of the American Statistical Association 96 (454) (2001) 717–729.
[9] P. Del Moral, Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications, Springer, New York, 2004.
[10] P. Del Moral, A. Doucet, A. Jasra, Sequential Monte Carlo for Bayesian Computation (with discussion). In Bayesian Statistics 8, Oxford University Press, 2006.
[11] P. Del Moral, A. Doucet, A. Jasrau, Sequential monte carlo samplers, Journal of the Royal Statistical Society B 68 (3) (2006) 411–436.
[12] M. Dorobantu, B. Engquist, Wavelet-based numerical homogenization, SIAM Journal of Numerical Analysis 35 (2) (1998) 540.
[13] P. Dostert, Y. Efendiev, T.Y. Hou, W. Luo, Coarse-gradient Langevin algorithms for dynamic data integration and uncertainty quantification, Journal of Computational Physics 217 (1) (2006) 123–142.
[14] A. Doucet, M. Briers, S. Senecal, Efficient block sampling strategies for sequential monte carlo, Journal of Computational and Graphical Statistics 15 (3) (2006) 693–711.

[15] A. Doucet, J.F.G. de Freitas, N.J. Gordon (Eds.), Sequential Monte Carlo Methods in Practice, Springer, New York, 2001.
[16] W. E, B. Engquist, The heterogeneous multi-scale methods, Communication in Mathematical Science 1 (2003) 87.
[17] AF. Emery, AV. Nenarokomov, TD. design Fadale, Uncertainties in parameter estimation: the optimal experiment, International Journal of Heat and Mass Transfer 43 (2000) 3331–3339.
[18] HW. Engl, M. Hanke, A. Neubauer, Regularization of Inverse Problems, Kluwer Academic Publishing, 1996.
[19] TD. Fadale, AV. Nenarokomov, AF. Emery, Uncertainties in parameter estimation: the inverse problem, International Journal of Heat and Mass Transfer 38 (3) (1995) 511–518.
[20] M. Fatemi, J.F. Greenleaf, Ultrasound-stimulated vibro-acoustic spectrography, Science 280 (5360) (1998) 82–85.
[21] M. Ferreira, H. Lee, Multiscale Modeling – A Bayesian Perspective, Springer Series in Statistics, Spinger, 2007.
[22] M.N. Fienen, P.K. Kitanidis, D. Watson, P. Jardine, An application of bayesian inverse methods to vertical deconvolution of hydraulic conductivity in a heterogeneous aquifer at oak ridge national laboratory, Mathematical Geology 36 (1) (2004) 101–126.
[23] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin, Bayesian Data Analysis, second ed., Chapman & Hall/CRC, 2003.
[24] R.E. Glaser, G. Johannesson, S. Sengupta, B. Kosovic, S. Carle, Stochastic engine final report: applying markov chain monte carlo methods with importance sampling to large-scale data-driven simulation, Technical Report, Lawrence Livermore National Lab., Livermore, CA, 2004.
[25] P.J. Green, Reversible jump markov chain monte carlo computation and bayesian model determination 82 (4) (1995) 711–732.
[26] C.W. Groetsch, Inverse Problems in the Mathematical Sciences, Vieweg, 1993.
[27] B.K. Hegstad, H. Omre, Uncertainty in production forecasts based on well observations seismic data and production history, Spe Journal 6 (4) (2001) 409–424.
[28] D. Higdon, Space and space-time modeling using process convolutions, in: Quantitative Methods for Current Environmental Issues, Springer-Verlag, 2002, pp. 37–56.
[29] D. Higdon, H. Lee, Z.X. Bi, A bayesian approach to characterizing uncertainty in inverse problems using coarse and fine-scale information, IEEE Transactions on Signal Processing 50 (2) (2002) 389–399.
[30] D. Higdon, H. Lee, C. Holloman, Markov chain monte carlo-based approaches for inference in computationally intensive inverse problems, in: Bayesian Statistics 7, Oxford University Press, 2003.
[31] D. Higdon, B. Nakhleb, J. Gattiker, B. Williams, A Bayesian calibration approach to the thermal problem, Computer Methods in Applied Mechanics and Engineering 197 (29–32) (2008) 2431–2441.
[32] C.H. Holloman, H. Lee, D. Higdon, Multi-resolution genetic algorithms and markov chain monte carlo, Journal of Computational and Graphical Statistics (2006) 861–879.
[33] T.Y. Hou, X.-H. Wu, A multiscale finite element method for elliptic problems in composite materials and porous media, Journal of Computational Physics 134 (1997) 169.
[34] T.J.R. Hughes, G.R. Feijóo, L. Mazzei, J.-B. Quincy, The variational multiscale method: a paradigm for computational mechanics, Computer Methods in Applied Mechanics and Engineering 166 (1–2) (1998) 3.
[35] W. Jefferys, J. Berger, Ockhams razor and bayesian analysis, American Science 80 (1992) 64–72.
[36] G Johannesson, R.E. Glaser, C.L. Lee, J.J. Nitao, W.G. Hanley. Multi-resolution markov-chain-monte carlo approach for system identification with an application to finite-element models, Technical Report, Lawrence Livermore National Lab., Livermore, CA, 2005.
[37] J.P. Kaipio, E. Somersalo, Computational and Statistical Methods for Inverse Problems, Springer-Verlag, New York, 2005.
[38] M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models, Journal of the Royal Statistical Society Series B – Statistical Methodology 63 (2001) 425–450.
[39] I.G. Kevrekidis, C.W. Gear, J.M. Hyman, P.G. Kevrekidis, O. Runborg, K. Theodoropoulos, Equation-free multiscale computation: enabling microscopic simulators to perform system-level tasks, Communications in Mathematical Sciences 1 (4) (2003) 715–762.
[40] G.S. Kimeldorf, G. Wahba, A correspondence between bayesian estimation on stochastic processes and smoothing by splines, Annals of Mathematical Statistics 41 (1971) 495–502.
[41] P.K. Kitanidis, Parameter uncertainty in estimation of spatial functions – bayesian-analysis, Water Resources Research 22 (4) (1986) 499–507.
[42] P.S. Koutsourelakis, Stochastic upscaling in solid mechanics: an excercise in machine learning, Journal of Computational Physics 226 (10) (2007) 301–325.
[43] H. Lee, D. Higdon, Z. Bi, M. Ferreira, M. West, Markov random field models for high-dimensional parameters in simulations of fluid flow in porous media, Technometrics 44 (3) (2002) 230–241.
[44] S.H. Lee, A. Malallah, A. Datta-Gupta, D. Higdon, Multiscale data integration using markov random fields, Spe Reservoir Evaluation and Engineering 5 (1) (2002) 68–78.
[45] M.S. Lewicki, T.J. Sejnowski, Learning overcomplete representations, Neural Computation 12 (2) (2000) 337–365.
[46] F. Liang, M. Liao, M. Mukherjee, S. nad West, Nonparametric bayesian kernel models, Technical Report, ISDS Discussion Paper, Duke University, 2006.
[47] F. Liu, M.J. Bayarri, J.O. Berger, R. Paulo, J. Sacks, A bayesian analysis of the thermal challenge problem, Computer Methods in Applied Mechanics and Engineering 197 (29–32) (2008) 2457–2466.
[48] J.S. Liu, Monte Carlo Strategies in Scientific Computing, Springer Series in Statistics, Springer, 2001.
[49] J.S. Liu, C. Sabatti, Generalized Gibbs sampler and multigrid monte carlo for bayesian computation, Biometrika 87 (2) (2000) 353–369.
[50] S.N. MacEachern, M. Clyde, J.S. Liu, Sequential importance sampling for nonparametric Bayes models: the next generation, The Canadian Journal of Statistics/La Revue Canadienne de Statistique 27 (2) (1998) 251–267.
[51] I. Murray, Z. Ghahramani, A note on the evidence and bayesian Occams razor, Technical Report, Gatsby Unit Technical Report GCNU-TR 2005-003, 2005.
[52] NSF Blue Ribbon Panel on Simulation-Based Engineering Science – revolutionizing engineering through simulation, Technical Report, National Science Foundation, 2006.
[53] C.E. Rasmussen, Z. Ghahramani, Occam's razor, Neural Information Processing Systems 13 (2001) 294–300.
[54] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, second ed., Springer, New York, 2004.
[55] G. Sangali, Capturing small scales in elliptic problems using a residual-free bubbles finite element method, SIAM Journal on Multiscale Modeling and Simulation 1 (3) (2003) 485.
[56] D.M. Schmidt, J.S. George, C.C. Wood, Bayesian inference applied to the electromagnetic inverse problem, Human Brain Mapping 7 (3) (1999) 195–212.
[57] J.C. Simo, T.J.R. Hughes, Computational Inelasticity, Interdisciplinary Applied Mathematics, Springer, 2000.
[58] A.N. Tikhonov, On the solution of incorrectly put problems and the regularization method, Academic Science in USSR Siberian Branch (1963) 261–265.
[59] A.N. Tikhonov, V.Y. Arsenin, Solution of Ill-posed Problems, John Wiley & Sons, 1977.
[60] M.E. Tipping, Sparse bayesian learning and the relevance vector machine, Journal of Machine Learning Research 1 (2001) 211–244.
[61] S. Torquato, Random Heterogeneous Materials, Springer-Verlag, 2002.
[62] B. Velamur Asokan, N. Zabaras, A stochastic variational multiscale method for diffusion in heterogeneous random media, Journal of Computational Physics 218 (2006) 654–676.
[63] J. Vermaak, S.J. Godsill, A. Doucet, Sequential bayesian kernel regression, Advances in Neural Information Processing Systems, 16, MIT Press, 2003.
[64] C. Vogel, Computational methods for inverse problems frontiers in applied mathematics series, SIAM, 2002.
[65] J. Wang, N. Zabaras, Permeability estimation using a hierarchical markov tree (hmt) model, in: Seventh World Congress on Computational Mechanics, Los Angeles, CA, July 16–22, 2006.
[66] J.B. Wang, N. Zabaras, A bayesian inference approach to the inverse heat conduction problem, International Journal of Heat and Mass Transfer 47 (17–18) (2004) 3927–3941.

[67] J.B. Wang, N. Zabaras, Hierarchical bayesian models for inverse problems in heat conduction, Inverse Problems 21 (1) (2005) 183–206.
[68] J.B. Wang, N. Zabaras, A markov random field model of contamination source identification in porous media flow, International Journal of Heat and Mass Transfer 49 (5–6) (2006) 939–950.
[69] I.S. Weir, Fully bayesian reconstructions from single-photon emission computed tomography data, Journal of the American Statistical Association 92 (437) (1997) 49–60.
[70] H.P. Wynn, P.J. Brown, C. Anderson, J.C. Rougier, P.J. Diggle, M. Goldstein, W.S. Kendall, P. Craig, K. Beven, K. Campbell, M.D. Mckay, P. Challenor, R.M. Cooke, N.A. Higgins, J.A. Jones, J.P.C. Kleijnen, W. Notz, T. Santner, B. Williams, J. Lehman, A. Saltelli, N. Shephard, H. Tjelmeland, M.C. Kennedy, A. O'Hagan, Bayesian calibration of computer models – discussion, Journal of the Royal Statistical Society Series B – Statistical Methodology 63 (2001) 450–464.