Safety, Reliability and Risk of Structures, Infrastructures and Engineering Systems – Furuta, Frangopol & Shinozuka (eds)
© 2010 Taylor & Francis Group, London, ISBN 978-0-415-47557-0

A multi-resolution Bayesian framework for the identification of spatial variability of material properties

P.S. Koutsourelakis
School of Civil and Environmental Engineering, Cornell University, USA

ABSTRACT: This paper proposes a hierarchical, multi-resolution framework for the identification of model parameters and their spatial 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 is based on a novel, adaptive Sequential Monte Carlo scheme which is directly parallelizable 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.

1 INTRODUCTION

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

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

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 $f(\boldsymbol{x})$, $\boldsymbol{x} \in \mathcal{D}$. 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 $\boldsymbol{x}_i \in \mathcal{D}$.

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. On the other hand, 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 (Lee, Higdon, Bi, Ferreira, and West 2002, Kitanidis 1986, Liu, Bayarri, Berger, Paulo, and Sacks 2008).

The accuracy of the predictions of computational models is greatly influenced by the the multiscale nature of property variations and a lot of research efforts have been devoted to the development of scalable, black box simulators that provide the coarse-scale solution while capturing the effect of fine-scale fluctuations (Dolbow, Khaleel, and Mitchell 2004). The multiscale analysis of such systems inherently assumes that the complete, fine-scale variation of various properties (or model parameters in general) is known. This assumption limits the applicability of these frameworks since it is usually not possible to experimentally determine the complete structure of the medium of interest at the finest scale. More often than not, what is experimentally available and accessible, are measurements of the response of these systems under prescribed input or excitation, at spatial scales much coarser than those of the property variations.

This limited and noisy information naturally introduces a lot of uncertainty and necessitates viewing

the property variation as a random field whose statistical properties must be consistent with the available data. 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 (Lee, Higdon, Bi, Ferreira, and West 2002). 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 exorbitant number of calls to the forward simulator in order to achieve convergence.

In the present paper we adopt a nonparametric, Bayesian model which is independent of the grid of the forward solver and is reminiscent of nonparametric 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

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. 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.

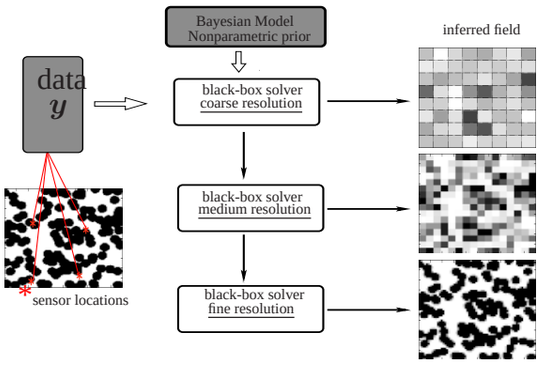


Figure 1. Hierarchy of solvers operating on different resolutions

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 (Figure 1).

In the remainder of this 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 2.3) the proposed inference techniques for the determination of the posterior.

2.1 Likelihood Specification

Let $\boldsymbol{F}^r = \{F_i^r\} : \mathcal{G} \rightarrow \mathcal{E}$ denote the vector-valued mapping implied by the forward model (operating at resolution r), which given $f(x) \in \mathcal{G}$ (Equation (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 Equation (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:

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

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

$$p_r(y_i \mid f(x), \sigma_r) \propto \frac{1}{\sigma_r} \exp\{-\frac{1}{2} \frac{\left(y_i - F_i^{(r)}(f(\boldsymbol{x}))\right)^2}{\sigma_r^2}\}$$

and

$$p_r(\boldsymbol{y} \mid f(x), \sigma_r) \propto \frac{1}{\sigma_r^n} \exp\{-\frac{1}{2\sigma_r^2} \sum_{i=1}^n \left(y_i - F_i^{(r)}(f(\boldsymbol{x}))\right)^2\} \tag{3}$$

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 Equation (3) further simplifying the likelihood:

$$L_r(f(x)) = p(\boldsymbol{y} \mid f(x)) \propto \Gamma(a + n/2) / \left(b + \tfrac{1}{2} \sum_{i=1}^n \left(y_i - F_i^{(r)}(f(\boldsymbol{x})\right)^2\right)^{a+n/2} \tag{4}$$

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

2.2 Prior Specification

The most critical component involves the prior specification for the unknown material properties as represented by $f(x)$. In existing Bayesian (Wang and Zabaras 2005), 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 *over-fitting*. 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 a discretized extension of a convolution representation of a non-stationary Gaussian process.

$$f(\boldsymbol{x}) = a_0 + \sum_{j=1}^{k} a_j K_j(\boldsymbol{x}; \boldsymbol{x}_j, \tau_j) \quad x \in D \tag{5}$$

where Gaussian isotropic kernels were used:

$$K(\boldsymbol{x}; \boldsymbol{x}_j, \tau_j) = \exp\{-\tau_j \parallel \boldsymbol{x} - \boldsymbol{x}_j \parallel^2\} \tag{6}$$

The parameters τ_j directly correspond to the scale of variability of $f(\boldsymbol{x})$. It should also be noted that other functional forms (e.g. discontinuous) for the kernels K_j can also be used on their own or in combinations

to enrich the expressivity of the expansion in Equation (5).

Such representations can be viewed as a radial basis network as in (Tipping 2001). Furthermore by interpreting the kernels as basis functions, Equation (5) it can be seen as an extension of the the representer theorem of Kimeldorf and Wahba (Kimeldorf and Wahba 1971). Overcomplete representations as in Equation (5) 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 (Lewicki and Sejnowski 2000).

An important improvement of the proposed formulation 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 Equation (5)) 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 (Equation (2), Figure 1)) as it will be illustrated in the sequence.

Due to space restrictions we briefly summarize the parameters of the model and the prior distributions used:

- k: the number of kernel functions. We employed a Poisson distribution $p(k \mid \lambda) = e^{-\lambda}\frac{\lambda^k}{k!}$ with an exponential hyper-prior for the hyper-parameter λ.
- $\{a_j\}_{j=1}^k$, the coefficients of the expansion in Equation (5). A multivariate normal $N(\mathbf{0}, \sigma_a^2\ \boldsymbol{I}_{k+1})$ was used and a inverse-Gamma hyper-prior for the hyper-parameter σ_a^2.
- $\{\tau_j\}_{j=1}^k$ the precision parameters of each kernel which pertain to the scale of the unknown field(s). We employed independent Gamma priors:

$$p(\{\tau_j\}_{j=1}^k \mid k,\ a_\tau, b_\tau) = \prod_{j=1}^{k} \frac{b_\tau^{a_\tau}}{\Gamma(a_\tau)} \tau_j^{a_\tau - 1}\ \exp(-b_\tau \tau_j) \tag{7}$$

  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.
- $\{\boldsymbol{x}_j\}_{j=1}^k$ the locations of the kernels which are points in $\mathcal{D}$ for which a uniform distribution in $\mathcal{D}$ was used.

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 Equation (5) the dimension of $\boldsymbol{\theta}_k$ is $(k+1) + k + 2k = 2 + 4k$. Based on the aforementioned equations, the complete prior model is given by:

$$\begin{aligned} &p(\boldsymbol{\theta} \mid s,\ a_\tau, a_\mu,\ a_0, b_0) = \tfrac{1}{(s+1)^{k+1}} \\ &\times \textstyle\prod_{j=1}^k \tfrac{\Gamma(a_\tau+1)}{\Gamma(a_\tau)}\ \tfrac{a_\tau^{a_\tau}}{\tau_j^{(a_\tau-1)}}\ \tfrac{1}{a_\mu}\ \tfrac{1}{(a_\tau \tau_j + a_\mu^{-1})^{(a_\tau+1)}} \\ &\times \tfrac{1}{(2\pi)^{(k+1)/2}} \tfrac{\Gamma(a_0 + \frac{k+1}{2})}{\left(b_0 + \frac{1}{2}\sum_{j=0}^k a_j^2\right)^{a_0 + (k+1)/2}} \times \tfrac{1}{|\mathcal{D}|^k} \end{aligned} \tag{8}$$

The combination of the prior $p(\boldsymbol{\theta})$ with the likelihood $L_r(\boldsymbol{\theta})$ (Equation (4)) corresponding to a forward solver operating on resolution r, give rise to the *posterior* density $\pi_r(\boldsymbol{\theta})$ which is proportional to:

$$\pi_r(\boldsymbol{\theta}) = p_r(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto L_r(\boldsymbol{\theta})\ p(\boldsymbol{\theta}) \tag{9}$$

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, Equation (9) 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 Equation (3) 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. 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.

2.3 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 (Higdon, Lee, and Bi 2002, Ferreira and Lee 2007). 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 (Liu 2001), the rate of convergence can be extremely slow and require a lot of likelihood evaluations and calls to the black-box solver. Particularly 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 (Doucet, de Freitas, and Gordon 2001)). As with Markov Chain Monte Carlo methods (MCMC), the target distribution(s) need only be known up to a constant. 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 (Del Moral, Doucet, and Jasrau 2006, Del Moral, Doucet, and Jasra 2006). 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 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:

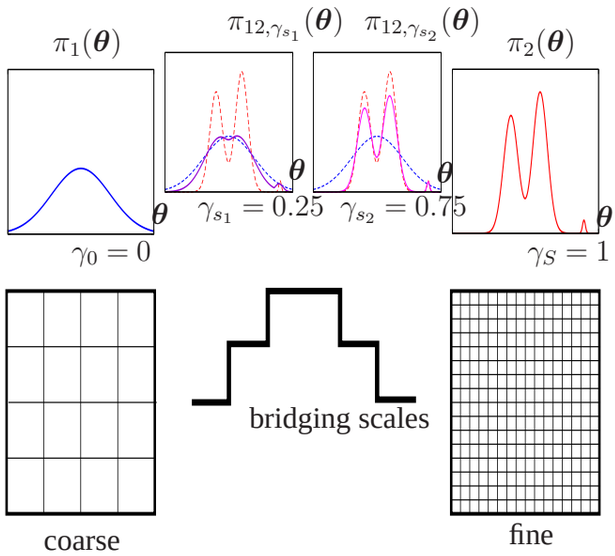


Figure 2. Illustration of bridging densities as defined in Equation (12) between posterior distributions $\pi_1(\boldsymbol{\theta})$, $\pi_2(\boldsymbol{\theta})$ corresponding to different resolutions of the governing PDEs. These allow for accurate and computationally efficient transmission of the inferences made to finer scales.

$$\pi_r(\boldsymbol{\theta}) \approx \sum_{i=1}^{N} W_r^{(i)}\, \delta_{\boldsymbol{\theta}_r^{(i)}}(\boldsymbol{\theta}) \tag{10}$$

where $W_r^{(i)} = w_r^{(i)} / \sum_{i=1}^N w_r^{(i)}$ are the normalized weights and $\delta_{\boldsymbol{\theta}_r^{(i)}}(.)$ is the Dirac function centered at $\boldsymbol{\theta}_r^{(i)}$. Furthermore, for any function $h(\boldsymbol{\theta})$ which is π_r-integrable (Del Moral 2004):

$$\sum_{i=1}^{N} W_r^{(i)}\, h(\boldsymbol{\theta}_r^{(i)}) \rightarrow \int h(\boldsymbol{\theta})\, \pi_r(\boldsymbol{\theta})\, d\boldsymbol{\theta} \quad \text{almost surely} \tag{11}$$

The goal is to obtain samples from each of the posterior distributions in Equation (9) 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} (Figure 2) and discuss the inferential transitions. Let $\pi_{12,\gamma}(\boldsymbol{\theta})$ denote a sequence of artificial, auxiliary distributions defined as follows ($\gamma \in [0,1]$)):

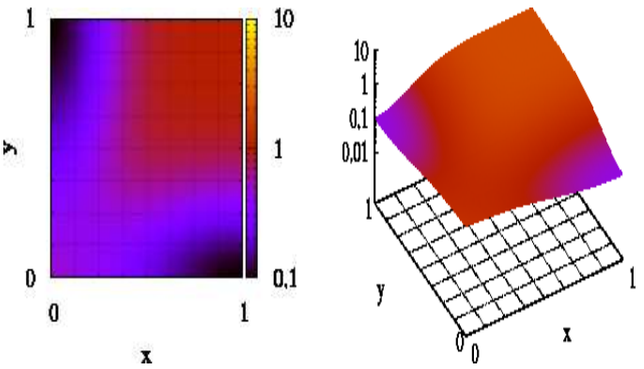
$$\pi_{12,\gamma}(\boldsymbol{\theta}) = \pi_1^{(1-\gamma)}(\boldsymbol{\theta})\ \pi_2^{\gamma}(\boldsymbol{\theta}) = L_1^{(1-\gamma)}(\boldsymbol{\theta})\ L_2^{\gamma}(\boldsymbol{\theta})\ p(\boldsymbol{\theta}) \tag{12}$$

where γ plays the role of *reciprocal temperature*. Trivially for $\gamma = 0$ we recover π_1 and for $\gamma = 1$, π_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. We have developed an adaptive SMC scheme (summarized in Table 1) where the necessary number of intermediate distributions is determined automatically.

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 $ESS_s \leq ESS_{min}$ then resample.
4. *Rejuvenate*: Use an MCMC kernel $P_s(.,.)$ that leaves π_{12,γ_s} invariant to perturb each particle $\boldsymbol{\theta}_{s-1}^{(i)} \rightarrow \boldsymbol{\theta}_s^{(i)}$
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 Equations (10), (11).
6. If $\gamma_s < 1$ then set $s = s + 1$ and goto to step 2. Otherwise stop.

Table 1. Basic steps of the *Adaptive* SMC algorithm proposed



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

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

3 NUMERICAL EXAMPLE

The method proposed is illustrated in a problem 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 (Simo and Hughes 2000). The field of interest in all the problems examined was the yield stress $\sigma_{yield}(\boldsymbol{x})$ which was assumed to vary spatially (Young's modulus $E = 1000$ and Poisson's ratio $\nu = 0.3$ were assumed known. 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.

The adaptive SMC scheme (Table 1) with $N = 100$ particles was employed in the examples presented with $\zeta = 0.95$ and $ESS_{min} = N/2$.

In order to generate data $\boldsymbol{y}$, it was assumed that the yield stress varied as follows (Figure 3):

$$\log \sigma_{yield}(\boldsymbol{x}) = -e^{-10\ x^2 - 2\ (y-1)^2} - e^{-2\ (x-1)^2 - 10\ y^2} \tag{13}$$

The nonlinear governing PDEs 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 $(0.125\ i,\ 0.125\ j)$, for $i = 1, \ldots, 8$ and $j = 0, \ldots, 8$ were recorded resulting in $n = 144$ data points (as in Figure 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 order to obtain sets of *observables* denoted by $\{y_i\}_{i=1}^n$ in our Bayesian model (Equation (2)). We note that in this example the scale of variability of the

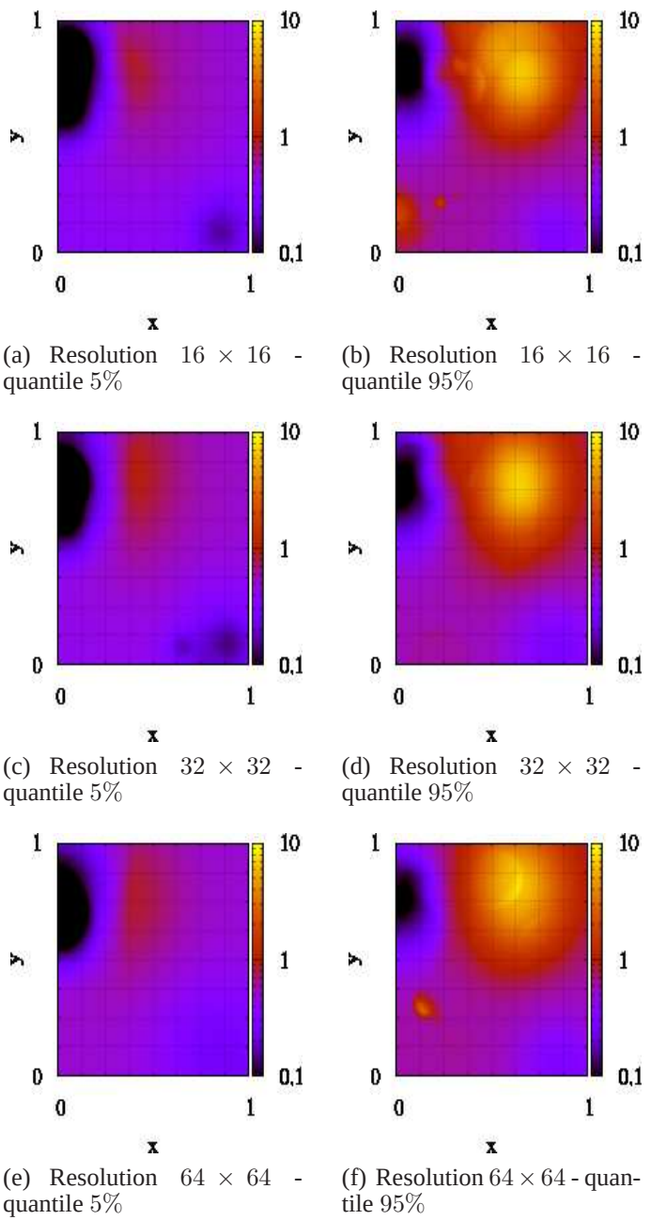
Solver Resolution	Degrees of Freedom	Normalized Computational Time (Actual in sec)
16×16	510	$\frac{1}{156}$ (0.55)
32×32	$2,046$	$\frac{1}{18}$ (4.8)
64×64	$8,190$	1 (86)

Table 2. Computational cost of different resolution solvers for Example A

unknown field $\sigma_{yield}(\boldsymbol{x})$ is *larger* than the scale of observations, i.e. the grid size where displacements were recorded.

Table 2 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. 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.

Figure 4 depicts the posterior quantiles obtained when a sequence of 3 solvers was used (16×16, 32×32 and 64×64). 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. Results on the computational effort are summarized in Table 3 which reports the *effective* computational cost at the various stages. The total number of calls was $6,265$ in contrast to $16,300$ that were required only when the finest solver was used (detailed results for this case are omitted due to space limitations) . Figure 5 depicts the posterior densities of the inferred model error standard deviations σ_r described in Equation (3). 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 $5\%\mu_A$.



(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%

Figure 4. Posterior quantiles at various solver resolutions for yield $\sigma_{yield}(\boldsymbol{x})$ Example A

Solver Resolution	Number of Bridging Distributions	Computational Effort (w.r.t. calls to 64×64 solver)
16×16	176	113
32×32	73	452
64×64	54	$5,700$
Total		$6,265$

Table 3. Computational cost for inferences. Note that the effective cost when using only the 64×64 solver was $16,300$

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

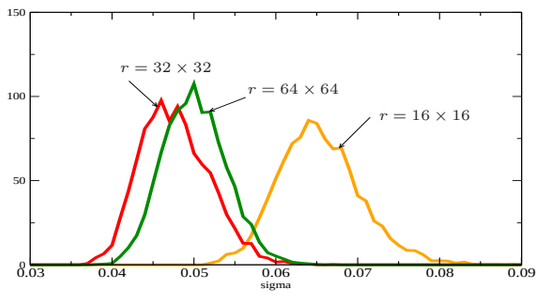


Figure 5. Posterior densities of model error st. deviations σ_r as in Equation (3). The values on x-axis have been divided by μ_A

based on SMC has been discussed which is embarrassingly parallelizable and well-suited for detecting multi-modal posterior distributions. The 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. 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. 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.

REFERENCES

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

Del Moral, P., Doucet, A., & Jasra, A. 2006. Sequential monte carlo for bayesian computation (with discussion). *Bayesian Statistics 8*. Oxford University Press.

Del Moral, P., Doucet, A., & Jasrau, A. 2006. Sequential Monte Carlo Samplers. *Journal of the Royal Statistical Society B* Vol. 68 (No. 3): 411–436.

Dolbow, J, Khaleel, M.A., & Mitchell, J 2004. Multiscale mathematics initiative: A roadmap. Technical report: Prepared for the U.S. Department of Energy under contract DE-AC05-76RL01830.

Doucet, A., de Freitas, J. F. G., & Gordon, N. J. (Eds.) 2001. *Sequential Monte Carlo Methods in Practice*. Springer New York.

Ferreira, M. & Lee, H. 2007. *Multiscale Modeling - A Bayesian Perspective*. Springer Series in Statistics. Spinger.

Higdon, D, Lee, H, & Bi, ZX 2002. A bayesian approach to characterizing uncertainty in inverse problems using coarse and fine-scale information. *IEEE TRANSACTIONS ON SIGNAL PROCESSING* Vol. 50 (No. 2): 389 – 399.

Kimeldorf, G.S. & Wahba, G. 1971. A correspondence between bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist. 41* 495–502.

Kitanidis, PK 1986. Parameter uncertainty in estimation of spatial functions - bayesian-analysis. *WATER RESOURCES RESEARCH* Vol. 22 (No. 4): 499 – 507.

Lee, H., Higdon, D., Bi, Z., Ferreira, M., & West, M. 2002. Markov random field models for high-dimensional parameters in simulations of fluid flow in porous media. Vol. 44 (No. 3): 230 – 241.

Lewicki, M.S. & Sejnowski, T. J. 2000. Learning overcomplete representations. *Neural Computation* Vol. 12 (No. 2): 337–365.

Liu, F, Bayarri, MJ, Berger, JO, Paulo, R, & Sacks, J 2008. A bayesian analysis of the thermal challenge problem. *COMPUTER METHODS IN APPLIED MECHANICS AND ENGINEERING* Vol. 197 (No. 29-32): 2457 – 2466.

Liu, J.S. 2001. *Monte Carlo Strategies in Scientific Computing*. Springer Series in Statistics. Springer.

Simo, J.C. & Hughes, T.J.R. 2000. *Computational Inelasticity*. Interdisciplinary Applied Mathematics. Springer.

Tipping, ME 2001. Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research, 1* 211–244.

Wang, JB & Zabaras, N 2005. Hierarchical bayesian models for inverse problems in heat conduction. *INVERSE PROBLEMS* Vol. 21 (No. 1): 183 – 206.