LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-TR-209485

Multi-Resolution Markov-Chain-Monte-Carlo Approach for System Identification with an Application to Finite-Element Models

G. Johannesson, R. E. Glaser, C. L. Lee, J. J. Nitao, W. G. Hanley

February 7, 2005

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

Multi-Resolution Markov-Chain-Monte-Carlo Approach for System Identification with an Application to Finite-Element Models

Gardar Johannesson Ronald E. Glaser Christopher L. Lee
John J. Nitao William G. Hanley

February 2, 2005

Lawrence Livermore National Laboratory

UCRL-TR-209485

Abstract

Estimating unknown system configurations/parameters by combining system knowledge gained from a computer simulation model on one hand and from observed data on the other hand is challenging. An example of such *inverse* problem is detecting and localizing potential flaws or changes in a structure by using a finite-element model and measured vibration/displacement data. We propose a probabilistic approach based on Bayesian methodology. This approach does not only yield a single best-guess solution, but a *posterior* probability distribution over the parameter space. In addition, the Bayesian approach provides a natural framework to accommodate prior knowledge. A Markov chain Monte Carlo (MCMC) procedure is proposed to generate samples from the posterior distribution (an ensemble of likely system configurations given the data). The MCMC procedure proposed explores the parameter space at different resolutions (scales), resulting in a more robust and efficient procedure. The large-scale exploration steps are carried out using coarser-resolution finite-element models, yielding a considerable decrease in computational time, which can be a crucial for large finite-element models. An application is given using synthetic displacement data from a simple cantilever beam with MCMC exploration carried out at three different resolutions.

Contents

1 Introduction

The problem considered here is that of damage detection and localization in a structure using a finite-element model and measured vibration and/or displacement data. This problem belongs to a broader class of *inverse* problems where one attempts to identify system configuration(s) that are in agreement with given observed output data. For complex systems, where for example the observed output data is linked to the system via complicated and computationally expensive numerical forward simulation model, the inverse problem can be challenging. Not only can it be prohibitively difficult to deterministically search for a single optimal system configuration that is in good agreement with the data, but there might be a variety of distinct system configurations that can account for the output data. This has prompted the search for alternative inverse approaches for complex systems that not only naturally account for the multiple system configurations that match the data, but are also able to leverage available prior knowledge about the system of interest. Bayesian methodology (e.g., Bernardo & Smith, 1994) is one such approach that has gained increased attention in recent years with advances in computational technology; see, for example Gilks et al. (1996) for applications in various fields and Aines et al. (2002) and Newmark et al. (2002) for application in geology using the Stochastic Engine framework. The Bayesian framework provides a natural way to transfer probabilistically stated *prior* knowledge of the current system into *posterior* knowledge (a distribution) that reflects both the prior information and the observed data. In all except trivial cases, a Markov Chain Monte Carlo (MCMC) sampling algorithm (see e.g., Gilks et al., 1996) is used to generate a set of system configurations from the posterior distribution. All posterior inferences are then accomplished using this MCMC sample (i.e., mean, standard deviation, quantiles, etc.).

In our application, the forward model consists of a numerical finite-element model of a physical structure, along the lines of Glaser et al. (2003). The parameters of the system are the material properties of each element in the model, in our case the stiffness of each element. Given a stiffness configuration, the finite-element model can be used to predict structural vibrations/displacements behavior, which can then be compared to corresponding observed behavior.

There are number of challenges in constructing a MCMC algorithm to sample stiffness configurations from the posterior distribution. The two main challenges stem from the potentially large number of elements involved and the bimodal nature of each element; it can either be nominal (as expected) or faulted (damaged). Hence, the prior distribution for each element is bimodal, with one mode representing nominal stiffness values while the other mode represents "faulted" stiffness values. As such, the MCMC algorithm has to be able to vary freely (jump) between the two prior stiffness modes (nominal versus faulted). As the problem is also high dimensional (the number of elements), a typical local exploration, where the next state of the chain is simply a small perturbation of the current state, can yield a badly mixed posterior sample, requiring the generation of an unrealistically

large number of samples (i.e., computationally expensive model evaluations). More severely, a locally-exploring MCMC algorithm can ignore important parts of the sample space as it is "stuck" sampling another portion of the sample space. For example, if two elements are actually flawed in a finite-element model, it is possible that an MCMC algorithm would get stuck sampling configurations that only capture one of the flaws.

Our goal is to build an efficient MCMC sampling algorithm that is able to quickly explore the large parameter space and yield a well mixed and representative sample. The approach we take is to alternate between small and large scale exploration. In addition, we propose to perform the large scale exploration using coarser-resolution finite-element models in order to reduce computational time. This approach brings in additional complexity as the number of elements (i.e., the dimension of the parameter space) is changing. This is a significant deviation from "classical" MCMC algorithms where the parameter vector is of fixed dimension. However, recently such problems have been tackled with success, examples being mixture models with unknown number of components (Richardson & Green, 1997) and the change point problem when the number of change points is unknown (Green, 1995). MCMC involving models that change dimension are often referred to as *trans-dimensional MCMC* (Green, 2002).

This report is organized as follows: Section 2 gives a brief introduction to Bayesian methodology and MCMC with emphasis on how to improve MCMC sampling. Section 3 follows with the definition of a Bayesian finite-element model which captures the potential nominal/flawed bimodal behavior of each element and introduces finite-element modeling at multiple resolutions. Section 4 gives a general overview of Green's reversible-jump MCMC (RJMCMC) approach to multi-resolution MCMC sampling. Section 5 follows then with an adaption of the RJMCMC to the multi-resolution finite-element model of Section 3. An application is finally given to synthetic data derived for a simple cantilever beam in Section 6.

2 Bayesian Methodology and Markov Chain Monte Carlo (MCMC)

Bayesian inference methodology (e.g., Bernardo & Smith, 1994) differs from the more classical (frequentest) statistical inference approach in that not only are the observed data considered random variables, but also the parameters of the underlying model (or system). Both approaches specify a probability distribution for the observed data conditional on the parameters of the model,

$$L(\mathbf{x}) \equiv p(\mathbf{y} \,|\, \mathbf{x}),$$

which we refer to as the *likelihood*, where

$$\mathbf{y} = \text{vector of observed data and}$$
$$\mathbf{x} = \text{vector of model parameters,}$$

and $p(\mathbf{y} \,|\, \mathbf{x})$ denotes the conditional distribution of $\mathbf{y}$ given $\mathbf{x}$. The frequentest approach aims at maximizing the above likelihood with respect to the unknown parameters $\mathbf{x}$, which can be an ill-posed and a daunting task for large models. In the Bayesian framework the parameters are assumed random and given a *prior* distribution,

$$p(\mathbf{x}),$$

that summarizes, in a probabilistic way, our knowledge about the parameters of the model prior to observing the data. Then, Bayes' theorem yields the *posterior* distribution,

$$\pi(\mathbf{x}) \equiv p(\mathbf{x} \,|\, \mathbf{y}) \propto p(\mathbf{y} \,|\, \mathbf{x}) p(\mathbf{x})$$

which probabilistically summarizes our knowledge of the model, given the observed data. For large models, inference based upon the above posterior distribution is carried out via stochastic sampling techniques, where in particular the Markov Chain Monte Carlo (MCMC) sampling methodology has been shown to be successful (e.g., Gilks et al., 1996). The MCMC approach generates a sample from a Markov chain that has the posterior distribution $\pi(\mathbf{x})$ as its limiting distribution. One way to accomplish this is to use the *Metropolis-Hastings* (M-H) algorithm. The M-H algorithm relies on a clever, user-specified proposal distribution to construct the needed Markov chain. If $\mathbf{x}$ is the current state of the chain in question, then a new state $\mathbf{x}'$ is proposed by a random draw from the proposal distribution, say $q(\mathbf{x}' \,|\, \mathbf{x})$. The newly proposed state is then accepted as the new state of the chain with probability $\alpha(\mathbf{x}, \mathbf{x}') = \min(1, \rho(\mathbf{x}, \mathbf{x}'))$, where

$$\rho(\mathbf{x}, \mathbf{x}') \equiv \frac{\pi(\mathbf{x}') q(\mathbf{x} \,|\, \mathbf{x}')}{\pi(\mathbf{x}) q(\mathbf{x}' \,|\, \mathbf{x})}$$

is the acceptance ratio. Note, to evaluate ρ, the likelihood needs to be evaluated at the proposed $\mathbf{x}'$. Hence, if the likelihood relies on a computationally expensive forward model (i.e., a model that predicts the data given the parameters), it is highly desirable, and often crucial to perform as few model evaluations as possible when exploring the posterior distribution. However, finding a proposal distribution that both explores the parameter space effectively and yields accepted proposals at a good rate is difficult. Typically, most proposal distributions propose a new set of parameters by only modifying a small subset (often just a single parameter) of the current parameter configuration. For systems that are well-behaved (locally smooth), this approach yields only a small change in the likelihood, hence a high acceptance rate, but results in a chain that traces the parameter space in a local manner, using very small steps that result in a badly mixed sample. In addition,

such a local exploration of the parameter space can trap the MCMC sampler in an area surrounded by a "wall" of low likelihood which can only be escaped by "bolder" (more global) parameter proposals. Numerous methods have been proposed to circumvent this problem. One popular approach is to work with a sequence of likelihoods, say $L_0(\mathbf{x}), L_1(\mathbf{x}), \ldots$, where $L_0(\mathbf{x}) = L(\mathbf{x})$ and $L_i(\mathbf{x})$, $i = 1, 2, \ldots$, are smoother ("flatter") versions of $L(\mathbf{x})$, for example, created by "heating up" the original likelihood; $L_i(\mathbf{x}) = L(\mathbf{x})^{1/T_i}$, $1 < T_1 < T_2 < \ldots$ (see Liu, 2001, on tempering). Hence, proposals evaluated using the flatter (hotter) likelihoods yield smaller changes in the likelihood, hence higher acceptance rate, allowing the sampler to escape local modes. The flatter likelihoods can then be combined with more global, large-scale proposals to facilitate better exploration of the parameter space. As outlined above, heated or not, the likelihood has the same computational overhead (i.e., the same forward model). However, for a large-scale proposal distribution evaluated using a heated (i.e., less accurate) likelihood, one might be able to utilize a less accurate (coarser resolution), computationally faster forward model. It is exactly this approach that we shall investigate; combining less accurate likelihoods with coarser-resolution forward models to facilitate MCMC sampling.

3 Bayesian Finite-Element Structural Flaw Detection Model

Assume we have a finite-element model of a physical structure (or a mechanical system), for example, of a house, of a bridge of an airplane, etc. Further, assume we have a prior knowledge about the material and other properties of each element that comprises the finite-element model, given that the current condition of the structure is as expected, that is, nominal, without any faults. However, given observed data that can be linked to the current condition of the structure, we want to identify elements, if any, that deviate from nominal behavior due to changes in element properties (e.g., material stiffness). Hence, our goal is to assert, in a probabilistic way, if the structure behaves as expected (i.e., nominal), and if not try to localize where and how much the structure deviates from nominal behavior (i.e., identify potentially flawed elements).

Let

$$s_i = \text{the stiffness (Young's modulus) value of the } i\text{-th element}$$

in a finite-element model, $i = 1, \ldots, N$, where N is the total number of elements in the model. Before observing any informative data on the current status of the structure, we assume that the prior distribution of each of the s_i is given by,

$$\begin{aligned} &p_0(s_i) \text{ if the } i\text{-th element is nominal,} \\ &p_1(s_i) \text{ if the } i\text{-th element is potentially flawed.} \end{aligned} \tag{1}$$

Typically, the prior distribution of a nominal element can be specified as a narrow distribution centered at the expected nominal stiffness value of each element, while flawed stiffness values have a much wider distribution covering a broad range of different types of potential flaws. We take both $p_0(s_i)$ and $p_1(s_i)$ to be given by a shifted and scaled Beta distribution with density,

$$p_\ell(s) = \frac{\Gamma(\alpha_\ell + \beta_\ell)}{\Gamma(\alpha_\ell)\Gamma(\beta_\ell)} \frac{1}{s_\ell^+ - s_\ell^-} \left(\frac{s - s_\ell^-}{s_\ell^+ - s_\ell^-} \right)^{\alpha_\ell - 1} \left(1 - \frac{s - s_\ell^-}{s_\ell^+ - s_\ell^-} \right)^{\beta_\ell - 1}, \quad s_\ell^- \le s \le s_\ell^+, \tag{2}$$

where $(s_\ell^-, s_\ell^+, \alpha_\ell, \beta_\ell)$, $s_\ell^- < s_\ell^+$, are all known parameters, $\ell = 0, 1$. Figure 1 shows $p_0(s)$ and $p_1(s)$ in the case when

$$(s_0^-, s_0^+, \alpha_0, \beta_0) = (1.80\text{e+}11, 2.00\text{e+}e11, 5, 5) \quad \text{and}$$
$$(s_1^-, s_1^+, \alpha_1, \beta_1) = (1.00\text{e+}9, 1.85\text{e+}11, 3, 3),$$

which are the parameter values used in our application in Section 6.

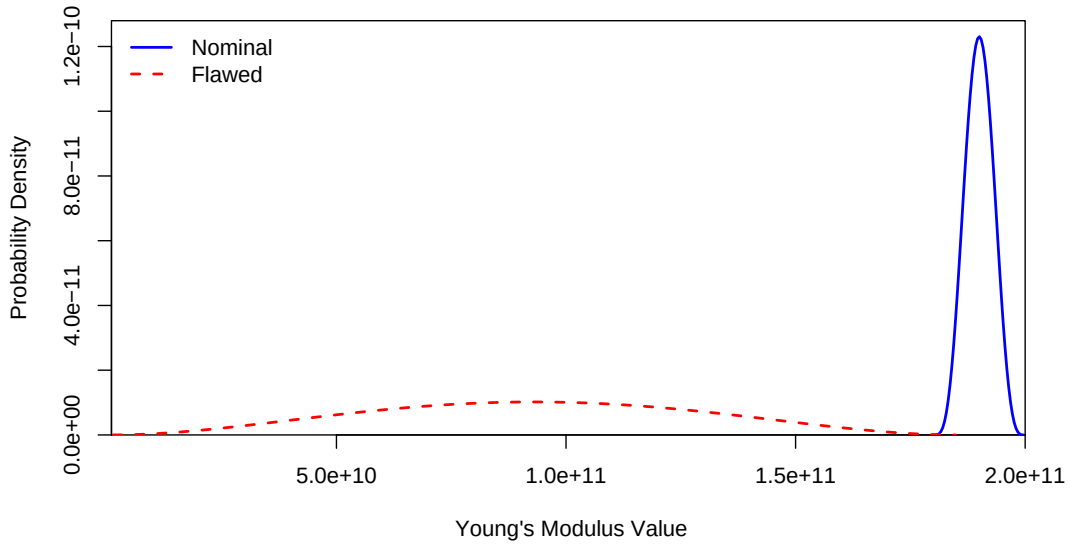

Figure 1: The prior distribution of Young's modulus values associated with nominal elements (right) and of potentially flawed elements (left).

The prior stiffness distribution (1) can also be written as

$$p(s_i \,|\, \ell_i) = (1 - \ell_i)p_0(s_i) + \ell_i p_1(s_i),$$

where

$$\ell_i = \begin{cases} 0 & \text{if the } i\text{-th element is nominal,} \\ 1 & \text{if the } i\text{-th element is potentially flawed.} \end{cases}$$

However, $\boldsymbol{\ell} \equiv \{\ell_i : i = 1, \ldots, N\}$ is not known and one of our goals is to perform inference on $\boldsymbol{\ell}$. We therefore treat $\boldsymbol{\ell}$ as a random variable and assume the following prior distribution,

$$p(\boldsymbol{\ell}) = \sum_{n=0}^{N} p(\boldsymbol{\ell} \,|\, n) p(n),$$

where

$$n \equiv \sum_{i=1}^{N} \ell_i = \text{the number of potentially flawed elements}$$

and $p(\boldsymbol{\ell} \,|\, n)$ is the prior distribution of $\boldsymbol{\ell}$ given there are n number of flaws. The discrete prior distribution for the number of flaws, $p(n)$, can be any discrete distribution on $0, \ldots, N$. Similarly, $p(\boldsymbol{\ell} \,|\, n)$ can be any proper distribution for the possible configurations of $\boldsymbol{\ell}$ that satisfy $n = \sum_i \ell_i$. We take $p(n)$ as the Binomial distribution with a known rate parameter λ;

$$p(n) = \binom{N}{n} \lambda^n (1-\lambda)^{N-n}, \quad n = 0, \ldots, N, \ \lambda \in [0, 1], \tag{3}$$

and $p(\boldsymbol{\ell} \,|\, n)$ as a uniform distribution on all possible configurations with n flaws;

$$p(\boldsymbol{\ell} \,|\, n) = 1 / \binom{N}{n},$$

where $\binom{N}{n}$ is the number of ways to have n flaws among N elements. Hence, in this case, the marginal prior distribution of $p(\boldsymbol{\ell})$ is simple $\boldsymbol{\ell}$ is $p(\ell) = \lambda^n (1-\lambda)^{N-n}$, which one could also derive by simply assuming that the elements are independently potentially flawed with prior probability λ (i.e., $\ell_i \sim$ Bernoulli(λ), independently, $i = 1, \ldots, N$).

The joint prior distribution of $\mathbf{s} \equiv \{s_i : i = 1, \ldots, N\}$ and $\boldsymbol{\ell}$ is then given by

$$p(\mathbf{s}, \boldsymbol{\ell}) = p(\mathbf{s} \,|\, \boldsymbol{\ell}) p(\boldsymbol{\ell}), \tag{4}$$

where $p(\mathbf{s} \,|\, \boldsymbol{\ell}) = \prod_{i=1}^{N} p(s_i \,|\, \ell_i)$ and $p(\boldsymbol{\ell}) = \sum_{n=0}^{N} p(\boldsymbol{\ell} \,|\, n) p(n)$.

As in the previous section, let $\mathbf{y}$ be the observed data and let

$$p(\mathbf{y} \,|\, \mathbf{s}) = \text{the probability distribution of } \mathbf{y}, \text{ conditional on } \mathbf{s}, \tag{5}$$

which we shall refer to as the *data model* or simply the *likelihood*. For example, $\mathbf{y}$ can be a set of nodal displacement data observed when applying a given force to the structure. In general, evaluating (5) requires running a computationally expensive forward model, the NIKE3D solver in our case, where the computational burden increases with N, the number of finite-elements in the model.

Given the data model (5) and the prior distribution (4), the posterior distribution of $\mathbf{s}$ and $\boldsymbol{\ell}$ is given by Bayes' theorem as

$$\pi(\mathbf{s}, \boldsymbol{\ell}) \equiv p(\mathbf{s}, \boldsymbol{\ell} \,|\, \mathbf{y}) \propto p(\mathbf{y} \,|\, \mathbf{s}) p(\mathbf{s} \,|\, \boldsymbol{\ell}) p(\boldsymbol{\ell}). \tag{6}$$

The posterior distribution above can then be used to answer a variety of questions related to the stiffness properties, including

- The posterior probability of the i-th element being potentially flawed, given by $\Pr(\ell_i = 1 \,|\, \mathbf{y})$
- The posterior probability of the total number of potentially flawed elements, as given by $p(n \,|\, \mathbf{y})$, and recall that $n = \sum_i \ell_i$.
- The posterior probability that the stiffness of the i-th element is lower than a given threshold, say s^o, given by $\Pr(s_i < s^o \,|\, \mathbf{y})$, or is within a given range, say from s^- to s^+, given by $\Pr(s^- < s_i < s^+ \,|\, \mathbf{y})$.

Due to the complexity of the model, the posterior distribution (6) is not available in closed form and we resort to MCMC to generate samples from the posterior distribution to use for inference. In doing so, there are two particular issues that need to be resolved. As N, the number of finite-elements, increases, the number of possible different configurations of $\boldsymbol{\ell}$ expands drastically, while at the same juncture, the time needed to evaluate the data model (5) increases due to the growing demands of a more complex (higher dimension) forward model. To be able to explore the parameter space sufficiently well as N increases, the Markov chain has to be both able to explore the parameter space in a local fashion (by proposing small changes in the parameter configuration) and in a global fashion (by proposing large changes in the parameter configuration) to be able to escape local modes of the posterior distribution and trace the parameter space more efficiently. To differentiate between two parameter configurations that are very similar (e.g., a small change in the spatial location of a single small structural flaw), a fine resolution finite-element model is needed. However, to differentiate between two parameter configurations that are significantly different (e.g., a large change in the spatial location of a single large structural flaw), a much coarser-resolution finite-element model is sufficient. Hence, to both reduce the computational time associated with evaluating the forward model and to facilitate better exploration of the parameter space, we propose using finite-element models at different resolutions of the same structure. The computationally expensive finer-resolutions models are used to explore finer details of the parameter space while the computationally efficient coarser-resolution models are used to explore the larger aspects of the structure.

Let $r = 0, \ldots, R$ index resolution, such that $r = 0$ corresponds to the finest-resolution finite-element model, and let N_r be the number of elements at resolution r. For the moment we shall focus on coarser-resolution models that are hierarchical in the sense that the resolution-r model is derived by collapsing (or aggregating)

one or more elements of the (finer) resolution-$(r-1)$ model into a single element. Now, let

$$\begin{aligned}\mathbf{s}_r &= \text{the stiffness configuration and}\\ \boldsymbol{\ell}_r &= \text{the 0/1, nominal/flawed, configuration}\end{aligned}$$

associated with resolution r, $r = 0, \ldots, R$. Our main goal is to perform inference at the finest resolution (i.e., about $\mathbf{s}_0$ and $\boldsymbol{\ell}_0$) and use the coarser-resolution models to guide and facilitate the MCMC sampling of $\mathbf{s}_0$ and $\boldsymbol{\ell}_0$. This requires not only traditional (single-resolution) proposal distributions within each resolution, but also proposal distributions between resolutions (trans-dimensional proposal distributions). Single-resolution (fixed-dimension) MCMC is well understood, while trans-dimensional MCMC, which we shall explore further in next section, has only recently gained attention.

4 Trans-Dimensional Models and Green's Reversible Jump MCMC (RJMCMC) Approach

Two approaches, or rather views, have been posted to construct a MCMC sampling algorithm for trans-dimensional models; the *reversible jump* approach (Green, 1995) and the *product-space* approach (Charlin & Chib, 1995). The former method constructs a Markov chain that jumps between models of different dimensions, while the second approach works with the joint distribution of all models under consideration, and hence a joint variable vector of fixed dimension (i.e., reducing the problem to a 'normal' MCMC of fixed dimension). It can be shown that the reversible jump approach is an important special case of the product-space approach (Besag, 1997, 2001). This connection between the the reversible jump and the product-space approach is then further investigated, and extended in Godsill (2001). In what follows we shall outline, in very general terms, Green's RJMCMC approach (Green, 1995); see also, for example, Andrieu et al. (2000, 2001), Waagepetersen & Sorensen (2001), and Green (2002) for overview, and Brooks et al. (2003) for construction of reversible jump proposal distributions and other issues associated with RJMCMC.

Our general notation is:

$\mathbf{y}$ is the observed data.

$\mathcal{M}_r$ is the r-th model under consideration; $r \in \{0, \ldots, R\}$ and note that R may be infinity.

$\mathbf{x}_r \in \mathcal{X}_r$ is the parameter vector associated with $\mathcal{M}_r$, and note that the $\mathbf{x}_r$'s are typically of different dimensions.

For the moment we shall assume that $\mathbf{x}_r$ is a continuous parameter vector. If $\mathbf{x}_r$ consists both of continuous and discrete components, as in our application where $\mathbf{x}_r = \{\mathbf{s}_r, \boldsymbol{\ell}_r\}$, the treatment of the discrete component will become obvious after the treatment of the continuous component which follows below.

Our model is given by:

$p_r(\mathbf{y} \,|\, \mathbf{x}_r)$, the conditional density of the observed data $\mathbf{y}$ given $\mathcal{M}_r$ configured by $\mathbf{x}_r$; i.e., the *data model* (the likelihood) using the forward model associated with $\mathcal{M}_r$.

$p(r, d\mathbf{x}_r) = p(d\mathbf{x}_r \,|\, r)p(r)$, the *prior distribution*, where $p(d\mathbf{x}_r \,|\, r)$ is the probability of $\mathbf{x}_r$ being in a infinitesimal set centered at $\mathbf{x}_r$ and we have used the shorthand notation, $p(\mathcal{M}_r) = p(r)$.

Our goal is to conduct inference based upon the posterior distribution of $(r, \mathbf{x}_r)$,

$$\pi(r, d\mathbf{x}_r) \equiv p(r, d\mathbf{x}_r \,|\, \mathbf{y}) = \pi_r(d\mathbf{x}_r)\pi(r)$$

via MCMC, where $\pi_r(d\mathbf{x}_r) \equiv p(d\mathbf{x}_r \,|\, r, \mathbf{y})$ and $\pi(r) \equiv p(r \,|\, \mathbf{y})$. In our application our focus is mainly on $\pi_0(d\mathbf{x}_0)$, the posterior distribution at the finest resolution. In other applications, for example model selection, the main interest is however on $\pi(r)$.

Denote by $(r, \mathbf{x}) = (r, \mathbf{x}_r)$ and $(r', \mathbf{x}') = (r', \mathbf{x}'_{r'})$ the current state and the next proposed state of the Markov chain, respectively. As in fixed-dimension MCMC the ingredients needed are:

$q(r', d\mathbf{x}' \,|\, r, \mathbf{x})$, the *proposal distribution* used to propose a new state $(r', \mathbf{x}')$ and

$\alpha(r, \mathbf{x}; r', \mathbf{x}')$, the *acceptance probability*, the probability of accepting the newly proposed state.

The relationship between the posterior distribution, the proposal distribution, and the acceptance ratio is given by (with a little abuse of notation)

$$\alpha(r, \mathbf{x}; r', \mathbf{x}') = \min\{1, \rho(r, \mathbf{x}; r', \mathbf{x}')\},$$

where

$$\rho(r, \mathbf{x}; r', \mathbf{x}') \equiv \frac{\pi(r', d\mathbf{x}')q(r, d\mathbf{x} \,|\, r', \mathbf{x}')}{\pi(r, d\mathbf{x})q(r', d\mathbf{x}' \,|\, r, \mathbf{x})}.$$

4.1 The Proposal Process

In fixed-dimension MCMC applications it is well defined what the reverse proposal move is for any given proposal. For example, if one proposes a new Markov chain state by simply adding a random number to one of the parameter component of the parameter vector $\mathbf{x}$, the reverse move is simply accomplished by subtracting the same

number from the resulting, proposed parameter. However, in trans-dimensional MCMC it is not as clear how one constructs the reverse proposal move for a move that changes the dimension of the parameter space. What is outlined below is a recipe proposed by Green (1995) and follows very much how one would implement MCMC in a computer program (fixed-dimensional MCMC or not).

The proposal $(r, \mathbf{x}) \to (r', \mathbf{x}')$ is constructed via:

(1) Select $r' \sim q_r(r' \,|\, \mathbf{x})$, and then

(2) let $\mathbf{x}' = g^x_{r\to r'}(\mathbf{x}, \mathbf{u})$, where $g^x_{r\to r'}$ is a deterministic function, $\mathbf{u} \in \mathcal{U}_{r\to r'}$, and $\mathbf{u} \sim q_{r\to r'}(d\mathbf{u} \,|\, \mathbf{x})$,

where in general q denotes a (proposal) probability distribution. Hence, one first proposes a new model index r' and then proposes $\mathbf{x}'$ as a function of the current value of the chain, $\mathbf{x}$, plus a random component, $\mathbf{u}$. The reverse proposal, $(r', \mathbf{x}') \to (r, \mathbf{x})$, is constructed in the same way:

(1) Select $r \sim q_{r'}(r \,|\, \mathbf{x}')$, and then

(2) let $\mathbf{x} = g^x_{r'\to r}(\mathbf{x}', \mathbf{u}')$; $\mathbf{u}' \in \mathcal{U}'_{r'\to r}$ and $\mathbf{u}' \sim q_{r'\to r}(d\mathbf{u}' \,|\, \mathbf{x}')$.

Now, to make this proposal process work both within the same model class (i.e., when $r' = r$) and between model classes (i.e., when $r' \neq r$ and $\mathbf{x}$ and $\mathbf{x}'$ are not of the same dimension) some care needs to be taken. We require that there is an one-to-one, differentiable mapping between $(\mathbf{x}, \mathbf{u})$ and $(\mathbf{x}', \mathbf{u}')$. This implies that there exist functions $g^u_{r\to r'}$ and $g^u_{r'\to r}$ such that

$$\mathbf{u}' = g^u_{r\to r'}(\mathbf{x}, \mathbf{u}) \text{ and } \mathbf{u} = g^u_{r'\to r}(\mathbf{x}', \mathbf{u}').$$

Hence:

$$\begin{aligned}(\mathbf{x}', \mathbf{u}') = g_{r\to r'}(\mathbf{x}, \mathbf{u}) \equiv \big(g^x_{r\to r'}(\mathbf{x}, \mathbf{u}), g^u_{r\to r'}(\mathbf{x}, \mathbf{u})\big) \text{ and} \\ (\mathbf{x}, \mathbf{u}) = g_{r'\to r}(\mathbf{x}', \mathbf{u}') \equiv \big(g^x_{r'\to r}(\mathbf{x}', \mathbf{u}'), g^u_{r'\to r}(\mathbf{x}', \mathbf{u}')\big),\end{aligned}$$

and in particular, $(\mathbf{x}, \mathbf{u}) = g_{r'\to r}(g_{r\to r'}(\mathbf{x}, \mathbf{u}))$. A necessary condition for this to hold is that

$$\dim(\mathbf{x}, \mathbf{u}) = \dim(\mathbf{x}', \mathbf{u}').$$

That is, the joint dimension of the parameters $\mathbf{x}$ and $\mathbf{u}$ has to be equal to the joint dimension of $\mathbf{x}'$ and $\mathbf{u}'$. (Recall we are assuming a continuous parameter spaces.)

Example. Assume we have two models, $\mathcal{M}_1$ and $\mathcal{M}_2$, where $\mathcal{M}_1$ is parameterized by x, a scalar, but $\mathcal{M}_2$ is parameterized by $\mathbf{x}' = (x'_1, x'_2)$, a vector of dimension 2. Say we propose to move from $\mathcal{M}_1$ to $\mathcal{M}_2$ by

$$\mathbf{x}' = (x'_1, x'_2) = g^x_{1\to 2}(x, u) = (x + u, x - u),$$

where $u \sim q(du)$ is a random perturbation. The reversible move is simply given by

$$x = g^x_{2\to 1}(\mathbf{x}', \mathbf{u}') = (x'_1 + x'_2)/2 \quad (\text{and } u = g^u_{2\to 1}(\mathbf{x}', \mathbf{u}') = (x'_1 - x'_2)/2),$$

and note that there is no random component involved (i.e., $\mathbf{u}'$ is of 'length' zero). We have that $\dim(\mathbf{x}, \mathbf{u}) = \dim(\mathbf{x}', \mathbf{u}')$, and equal to 2, as required.

Remark. The above presentation assumed that different proposals are only associated with the model index r. However, one might have more than one type of proposal move to jump from a parameter configuration of $\mathcal{M}_r$ to a parameter configuration of $\mathcal{M}_{r'}$. An index for different proposal types (types of moves) can be included in the parameter vector $\mathbf{u}$. That is, one would write

$$\mathbf{u} = (m, \mathbf{u}_m),$$

where m indexes move type and $\mathbf{u}_m$ is the parameter vector used to construct a proposal from $\mathcal{M}_r$ to $\mathcal{M}_{r'}$ of type m. In this case,

$$\mathbf{x}' = g^x_{r\to r'}(\mathbf{x}, \mathbf{u}) = g^{x(m)}_{r\to r'}(\mathbf{x}, \mathbf{u}_m)$$

and

$$\mathbf{u}' = (m, \mathbf{u}'_m) = g^u_{r\to r'}(\mathbf{x}, \mathbf{u}) = (m, g^{u(m)}_{r\to r'}(\mathbf{x}, \mathbf{u}_m)).$$

The proposal distributions $q_{r\to r'}$ consist then of two steps; proposing a move-type, $m \sim q_{r\to r'}(m \,|\, \mathbf{x})$, and then $\mathbf{u}_m \sim q^m_{r\to r'}(d\mathbf{u}_m \,|\, \mathbf{x})$. Similarly, we have a reverse proposal move, from $\mathcal{M}_{r'}$ to $\mathcal{M}_r$ using a proposal of type m. And as before, we assume that there is an one-to-one relationship between $(\mathbf{x}, \mathbf{u}_m)$ and $(\mathbf{x}', \mathbf{u}'_m)$ for each m.

4.2 The Acceptance Ratio

Assume that $(r, \mathbf{x}) \neq (r', \mathbf{x}')$; that is, there is a zero probability of proposing the next state of the Markov chain to be equal to the current state of the chain. The acceptance ratio can be written as

$$\begin{aligned}\rho(r, \mathbf{x}; r', \mathbf{x}') &= \frac{\pi(r', d\mathbf{x}')q_{r'\to r}(r \,|\, \mathbf{x}')q_{r'\to r}(d\mathbf{u}' \,|\, \mathbf{x}')}{\pi(r, d\mathbf{x})q_{r\to r'}(r' \,|\, \mathbf{x})q_{r\to r'}(d\mathbf{u} \,|\, \mathbf{x})} \\ &= \frac{\pi(r', \mathbf{x}')q_{r'\to r}(r \,|\, \mathbf{x}')q_{r'\to r}(\mathbf{u}' \,|\, \mathbf{x}')}{\pi(r, \mathbf{x})q_{r\to r'}(r' \,|\, \mathbf{x})q_{r\to r'}(\mathbf{u} \,|\, \mathbf{x})} \times \frac{\mu'_x(d\mathbf{x}')\mu'_u(d\mathbf{u}')}{\mu_x(d\mathbf{x})\mu_u(d\mathbf{u})},\end{aligned} \tag{7}$$

where μ and μ' are the underlying measures (i.e, $\pi(r, d\mathbf{x}) = \pi(r, \mathbf{x})\mu_x(d\mathbf{x})$, etc.). Using the (one-to-one) mapping from $(r, \mathbf{x}) \to (r', \mathbf{x}')$, given by $g_{r\to r'}$, the ratio of the measures can be written as

$$\frac{\mu'_x(dg^x_{r\to r'}(\mathbf{x}, \mathbf{u}))\mu'_u(dg^u_{r\to r'}(\mathbf{x}, \mathbf{u}))}{\mu_x(d\mathbf{x})\mu_u(d\mathbf{u})}.$$

Then, in the case when all the underlying measures are Lebesgue and $g_{r\to r'}$ is differentiable, the ratio above is given by

$$J_{r\to r'}(\mathbf{x},\mathbf{u};\mathbf{x}',\mathbf{u}') \equiv \left|\det \frac{\partial g_{r\to r'}(\mathbf{x},\mathbf{u})}{\partial(\mathbf{x},\mathbf{u})}\right|,$$

the Jacobian for the change-of-variables transformation $g_{r\to r'}$. Hence, in this case, the acceptance ratio is given by

$$\rho(r,\mathbf{x};r',\mathbf{x}') = \frac{\pi(r',\mathbf{x}')q_{r'\to r}(r\,|\,\mathbf{x}')q_{r'\to r}(\mathbf{u}'\,|\,\mathbf{x}')}{\pi(r,\mathbf{x})q_{r\to r'}(r'\,|\,\mathbf{x})q_{r\to r'}(\mathbf{u}\,|\,\mathbf{x})} \times J_{r\to r'}(\mathbf{x},\mathbf{u};\mathbf{x}',\mathbf{u}'). \tag{8}$$

Example (continued). The Jacobian for the transformation $(x_1', x_2') = (x+u, x-u)$ can be seen to be equal to -2. Hence, the proposal ratio $\rho_{1\to 2}(x,u;x_1',x_2')$ is given by,

$$\rho_{1\to 2}(x,u;x_1',x_2') = \frac{\pi_2(\mathbf{x}')}{\pi_1(x)q(u)} \times 2.$$

There are number of cases where the underlying measure for all the variables involved is Lebesgue. However, in our application, $\mathbf{x}_r = (\boldsymbol{\ell}_r, \mathbf{s}_r)$, where $\boldsymbol{\ell}_r = (\ell_1,\ldots,\ell_{N_r})$ and $\mathbf{s}_r = (s_1,\ldots,s_{N_r})$, and $\boldsymbol{\ell}_r$ is an N_r-dimensional discrete variable while $\mathbf{s}_r$ is an N_r-dimensional continuous positive variable. Hence, the underlying measure for $\boldsymbol{\ell}_r$ is a N_r-dimensional counting measure while the underlying measure for $\mathbf{s}_r$ is the Lebesgue measure. However, it is only for the continuous variables that one must deal with the ratios of the underlying measures in the proposal process. In the proposal process for the discrete component one only has to satisfy a one-to-one relationship in order to construct the reversible move.

5 The Finite-Element Model RJMCMC

Recall that at any given model resolution the prior distribution is (ignoring the resolution index r)

$$p(\mathbf{s},\boldsymbol{\ell}) = p(\mathbf{s}\,|\,\boldsymbol{\ell})p(\boldsymbol{\ell}),$$

where

- $p(\mathbf{s}\,|\,\boldsymbol{\ell}) = \prod_{i=1}^{N} p(s_i\,|\,\ell_i)$ where $p(s_i\,|\,\ell_i) = (1-\ell_i)p_0(s_i) + \ell_i p_1(s_i)$, $p_0(s_i)$ and $p_1(s_i)$ are both shifted and scaled Beta distributions; see (2) and Figure 1
- $p(\ell) = \sum_{n=0}^{N} p(\ell\,|\,n)p(n)$ where $n = \sum_{i=1}^{N}\ell_i$ = the number of flaws, $p(n)$ is Bin(λ) and $p(\ell\,|\,n) = 1/\binom{N}{n}$.

Currently, we assume that all the parameters associated with the likelihood (the data model) are known. However, in practice the likelihood might include unknown parameters, for example, a measurement+model error variance parameter, which can be included in the set of model parameters and sampled.

We shall now outline the different RJMCMC proposal moves and the current 'default' values of parameters associated with the proposal distributions. Note that in this initial study these parameters have not been 'tuned' in any way to yield optimal acceptance ratios.

5.1 Proposing What Move to Make Next

Assume that the Markov chain is currently at resolution r. The next type of proposal move is selected via two stage process:

1. With a user inputed probabilities ($q_r(r' \,|\, \mathbf{x})$) the chain either stays at its current resolution, or goes up or down one resolution. By default, the chain stays at its current resolution with probability 0.9 (else goes up or down with equal probability).

2. Given the next proposed resolution, r', the type of proposal move is selected among all the possible proposal moves available from resolution r to resolution r' (with given probabilities). By default, equal probabilities are assigned to each move within a resolution and between resolutions.

Hence, the multiplicative factor that contributes to $\rho(r, \mathbf{x}; r', \mathbf{x}')$ in (7) is

$$\frac{q_{r'}(r \,|\, \mathbf{x}')q_{r' \to r}(m \,|\, \mathbf{x}')}{q_r(r' \,|\, \mathbf{x})q_{r \to r'}(m \,|\, \mathbf{x})}.$$

We shall now outline the different proposals available within a resolution and between resolutions

5.2 Within-a-Resolution Proposals

In what follows, $q(X)$ denotes a generic proposal distribution for a random variable X that may depend on the current status of the Markov chain (i.e., on r and $\mathbf{x}$). We currently have four different types of proposals within a resolution.

Move 1: Birth/Death of a Flaw

This type of proposal either adds a flaw (birth) or removes a flaw (death) using the following procedure:

1. A 0/1 Bernoulli variable $a \sim q(a)$ is drawn with a given success probability and if $a = 1$, a birth of a flaw is proposed, else a death of a flaw. Currently, $\Pr(a = 1) = p(n+1)/(p(n-1)+p(n+1))$, where $p(n)$ is the prior distribution on the total number of flaws.

2. A location index $k \sim q(k \,|\, a)$ is selected randomly (uniformly) among nominal elements if $a = 1$ or among flawed elements if $a = 0$.

3. A new stiffness value, $v \sim q(dv \,|\, k, a)$, is generated in the nominal range, if $a = 0$, and in the flawed range if $a = 1$ and $s'_k = v$ and $\ell'_k = a$. We take $q(dv \,|\, k, a)$ as the nominal prior distribution of the k-th element if $a = 0$ or as the flawed prior distribution if $a = 1$.

Hence, in the notation of previous section,

$$\mathbf{u} = (a, k, v) \sim q(dv \,|\, a, k) q(k \,|\, a) q(a).$$

The random parameter $\mathbf{u}'$ of the reverse move is easily derived from $\mathbf{u}$ and $\mathbf{x}$ as $\mathbf{u}' = (a', k', v')$ where $a' = 0$ if $a = 1$, else $a' = 1$ if $a = 0$, $k' = k$, and $v' = s_k$.

Move 2: Relocate a Flaw

A flaw is relocated to a new location (keeping the same stiffness value), where a higher probability is given to a nearest neighbor relocation. The proposal process is as follows:

1. Draw an element location index $k_1 \sim q(k_1)$ among the flawed locations, where $q(k_1)$ is currently taken to be a uniform distribution.

2. Draw a 0/1 Bernoulli variable $l \sim q(l \,|\, k_1)$, where if $l = 1$, a relocation to a nominal nearest neighbor location is suggested, else if $l = 0$, to any nominal location. Currently, if k_1 has at least one nominal nearest neighbor, $\Pr(l = 1) = 0.9$.

3. Draw a relocation index, $k_2 \sim q(k_2 \,|\, l, k_1)$, where currently $q(k_2 \,|\, l, k_1)$ is either a uniform distribution over all nominal nearest-neighbor locations (if $l = 1$) or all nominal locations (if $l = 0$).

4. Swamp stiffness values; $s = s_{k_2}$, $s_{k_2} = s_{k_1}$, $s_{k_1} = s$.

Hence,

$$\mathbf{u} = (k_1, l, k_2) \sim q(k_1) q(l \,|\, k_1) q(k_2 \,|\, l, k).$$

The reverse move is trivial, with $\mathbf{u}' = (k'_1, l', k'_2)$ where $k'_1 = k_2$, $l' = l$, and $k'_2 = k_1$.

Move 3: Resize a Flaw

Change the stiffness value of a flawed element. The proposal process is:

1. Draw a element location index $k \sim q(k)$ among the flawed locations, where $q(k)$ is currently a uniform distribution over all flawed locations.

2. Draw a stiffness change value $\Delta s \sim q(d\Delta s \,|\, k)$ and let $s'_k = s_k + \Delta s$. The only constraint on the distribution $q(d\Delta s \,|\, k)$ is that the new stiffness value is still considered a potential flaw (i.e., $\ell'_k = \ell_k = 1$). In our current implementation,

$q(d\Delta s \,|\, k)$ is taken to be a shifted and scaled Beta distribution with range $(s_1^- - s, s_1^+ - s)$ and recall that $s_1^- \leq s \leq s_1^+$. The two shape parameters of the Beta distribution $q(d\Delta s \,|\, k)$, say α and β, are selected such to yield a mode at zero and constrained such that $\alpha + \beta = c \cdot (\alpha_1 + \beta_1)$, where α_1 and β_2 are the shape parameters of the prior distribution of flawed stiffness values. The scaling parameter c is taken to be equal to 2 be default (based on a small empirical study).

Hence, we have

$$\mathbf{u} = (k, \Delta s) \sim q(k)q(d\boldsymbol{\Delta} \,|\, k).$$

The reverse move has $\mathbf{u}' = (k', \Delta s')$, where $k' = k$ and $\Delta s = -\Delta s$.

Move 4: Resize a Nominal

Change the stiffness value of a nominal element. The proposal process is as in the case of resizing the stiffness of a flawed element. The range of the shifted and scaled Beta distribution involved is taken to be $(s_0^- - s, s_0^+ - s)$ and the two shape parameters, say α and β, are taken to yield a mode at zero and constrained such that $\alpha + \beta = c \cdot (\alpha_0 + \beta_0)$, where the scaling parameter $c = 2$ by default.

Other Possible Moves

Possible additional moves include updating potential variance parameters associated with the likelihood, particularly a scaling parameter associated with the size of the model error. Note that this move does not require a new evaluation of the forward model and is therefore 'inexpensive'.

5.3 Between Resolution Moves

To simplify notation (and to be in line with our final application), assume that each element at resolution r is derived by aggregating two elements at resolution $(r-1)$. We shall refer to the resulting aggregated coarse-resolution element as the parent and the two finer-resolution elements as the children (then again, the parent element at resolution r is a child of an element at resolution $(r+1)$, and so forth). Denote by s_p the stiffness and ℓ_p the status of a coarse-resolution parent element and denote by s_{c_1} and s_{c_2} the stiffness and ℓ_{c_1} and ℓ_{c_2} the status of its two children elements. We consider two different type of moves, based on a *weighted arithmetic mean* aggregation and a *weighted harmonic mean* aggregation, respectively. Both moves have deterministic fine-to-coarse-resolution proposal while the coarse-to-fine-resolution proposals involve random variables (the dimension-matching variables). We shall outline both type of moves, with slightly more emphasis on the trans-dimensional move based on weighted harmonic mean aggregation.

Weighted Harmonic Mean Aggregation (WHMA)

The deterministic fine-to-coarse-resolution proposal move is given by

$$s'_p = 1/(w_{c_1}/s_{c_1} + w_{c_2}/s_{c_2}), \qquad \text{(weighted harmonic mean)}$$

where the weights w_{c_1} and w_{c_2} sum to one. The parent element is considered potentially flawed, $\ell'_p = 1$, if at least one of the children is potentially flawed (that is, if $\ell_{c_1} + \ell_{c_2} > 0$).

The coarse-to-fine-resolution proposal move is given by

$$s'_{c_1} = s_p w_{c_1}(1 + u_p) \text{ and } s'_{c_2} = s_p w_{c_2}(1 + 1/u_p),$$

where $u_p \sim q(u_p)$, $u_p > 0$. Note that this yields

$$s_p = 1/(w_{c_1}/s'_{c_1} + w_{c_2}/s'_{c_2}) \text{ and } u_p = (w_{c_2}/s'_{c_2})/(w_{c_1}/s'_{c_1}).$$

The proposal distribution $q(u_p)$ is taken to be a shifted and scaled Beta distribution with parameters that depend on the status of the parent element (ℓ_p) and its Young's modulus value (s_p). There are two main cases:

The Parent Element is Nominal ($\ell_p = 0$). In this case both children are proposed to be nominal (i.e., $\ell'_{c_1} = \ell'_{c_2} = 0$). This is accomplished by specifying the range of the shifted and scaled Beta proposal distribution $q(u_p)$ such that both s'_{c_1} and s'_{c_2} are in the nominal prior range. By default, the mean of $q(u_p)$ is taken to be equal to the midpoint of the range, which only leaves one degree of freedom to fully specify the two Beta scaling parameters of $q(u_p)$, say α and β. The current default approach is to constrain $\alpha + \beta = \hat{\alpha} + \hat{\beta}$, where $\hat{\alpha}$ and $\hat{\beta}$ are the shape parameters of a Beta distribution approximation to the true distribution of the weighted harmonic mean of s'_{c_1} and s'_{c_2}, when both are assumed to be distributed according to the nominal prior distribution. The (multiplicative) contribution to the proposal ratio $\rho(r, \mathbf{x}; r', \mathbf{x}')$ of (8) is

$$\frac{q_{r' \to r}(\mathbf{u}' \,|\, \mathbf{x}')}{q_{r \to r'}(\mathbf{u} \,|\, \mathbf{x})} \times J_{r \to r'}(\mathbf{x}, \mathbf{u}; \mathbf{x}', \mathbf{u}') = \frac{1}{q(u_p)} \times \left(s_p w_{c_1} w_{c_2} (1 + 1/u_p)^2 \right).$$

The Parent Element is Potentially Flawed ($\ell_p = 1$). In this case either one or both children are proposed potentially flawed. The proposal process is:

1. Draw a 0/1 Bernoulli variable a with $\Pr(a = 1) = q(a = 1)$. By default, $q(a = 1) = q_2(s_p)/(q_1(s_p) + q_2(s_p))$, where $q_1(s_p)$ and $q_2(s_p)$ are Beta distribution approximations to the prior distributions of the weighted harmonic mean of s'_{c_1} and s'_{c_2} when either one of them or both of them are potentially flawed, respectively.
2. Then, according to the value of a:

i If $a = 1$, then both children are proposed to be potentially flawed ($\ell'_{c_1} = \ell'_{c_2} = 1$). The parameters of $q(u_p)$ are specified in a manner similar to the case when both children are proposed to be nominal, except now they are both potentially flawed. In this case the (multiplicative) contribution to the proposal ratio $\rho(r, \mathbf{x}; r', \mathbf{x}')$ of (8) is

$$\frac{1}{q(a=1)q(u_p)} \times \left(s_p w_{c_1} w_{c_2} (1 + 1/u_p)^2\right).$$

ii If $a = 0$, then one of the children is proposed potentially flawed (with equal probability assigned to each). Again, the parameters of $q(u_p)$ are specified in a manner similar to the case above, but with one of children assumed nominal and the other one potentially flawed. The contribution to the proposal ratio $\rho(r, \mathbf{x}; r', \mathbf{x}')$ is

$$\frac{1}{0.5q(a=0)q(u_p)} \times \left(s_p w_{c_1} w_{c_2} (1 + 1/u_p)^2\right).$$

Weighted Arithmetic Mean Aggregation (WAMA)

In this approach, the deterministic fine-to-coarse-resolution proposal move is given by

$$s'_p = w_{c_1} s_{c_1} + w_{c_2} s_{c_1} \qquad \text{(weighted arithmetic mean)}$$

where the weights w_{c_1} and w_{c_2} sum to one. As in the WHMA approach, the parent element is considered potentially flawed if at least one of the children is potentially flawed.

The coarse-to-fine-resolution proposal move is given by

$$s'_{c_1} = s_p + w_{c_1}^{-1} u_p \text{ and } s'_{c_2} = s_p - w_{c_2}^{-1} u_p,$$

where $u_p \sim q(u_p)$. As in the WHMA approach, the proposal distribution $q(u_p)$ is taken to be a shifted and scaled Beta distribution with parameters that depend on the status of the parent element, such that if:

The Parent Element is Nominal. Then both children are proposed to be nominal and the parameters of $q(u_p)$ are specified accordingly (and guided by the prior distribution). The (multiplicative) contribution to the proposal ratio $\rho(r, \mathbf{x}; r', \mathbf{x}')$ of (8) is

$$\frac{q_{r' \to r}(\mathbf{u}' \,|\, \mathbf{x}')}{q_{r \to r'}(\mathbf{u} \,|\, \mathbf{x})} \times J_{r \to r'}(\mathbf{x}, \mathbf{u}; \mathbf{x}', \mathbf{u}') = \frac{1}{q(u_p)} \times \left(\frac{1}{w_{c_1} w_{c_2}}\right).$$

The Parent Element is Potentially Flawed. Then either one or both children are proposed potentially flawed, as in the WHMA case. The proposal process is:

1. Draw a 0/1 Bernoulli variable a with $\Pr(a = 1) = q(a = 1)$ specified in a manner similar to the WHMA case.
2. Then, according to the value of a:
   i If $a = 1$, then both children are proposed to be potentially flawed and the parameters of $q(u_p)$ are specified accordingly. This yields

$$\frac{q_{r'\to r}(\mathbf{u}' \,|\, \mathbf{x}')}{q_{r\to r'}(\mathbf{u} \,|\, \mathbf{x})} \times J_{r\to r'}(\mathbf{x}, \mathbf{u}; \mathbf{x}', \mathbf{u}') = \frac{1}{q(a=1)q(u_p)} \times \left(\frac{1}{w_{c_1} w_{c_2}}\right),$$

   ii If $a = 0$, then one of the children is proposed potentially flawed (with equal probability) and the parameters of $q(u_p)$ specified accordingly. This yields

$$\frac{q_{r'\to r}(\mathbf{u}' \,|\, \mathbf{x}')}{q_{r\to r'}(\mathbf{u} \,|\, \mathbf{x})} \times J_{r\to r'}(\mathbf{x}, \mathbf{u}; \mathbf{x}', \mathbf{u}') = \frac{1}{0.5q(a=0)q(u_p)} \times \left(\frac{1}{w_{c_1} w_{c_2}}\right),$$

6 Application to Simple Cantilever Beam

Our application is to a finite-element model of a simple cantilever beam of length 100 with one of its end anchored to a wall, as in Glaser et al. (2003). The characteristics of the beam are modeled (at the finest resolution, resolution 0) by a finite-element model with 40 element of identical size. A coarser 20 element model (resolution 1) is formed by joining pairs of adjacent elements of the 40 element model, and similarly, a 10 element model (resolution 2) is formed by joining pairs of adjacent elements of the 20 element model. Hence, the 11 nodes of the resolution-2 10 element model are at locations (measured from the wall) 0, 10, 20, ..., 100, while for the resolution-1 20 element model the node locations are 0, 5, 10, ..., 100, and for the resolution-0 40 element model the node locations are 0, 2.5, 5, ..., 100. Figure 2 shows a schematic view of the beam at the three resolutions.

The Young's modulus value of the material used to construct the beam is specified to be 1.9e+11. However, the actual Young's modulus values associated with the (as-build, nominal) 40 elements of the resolution-0 model are not taken to be exactly equal to this nominal value, but rather are randomly scattered around 1.9e+11 (which can be thought to be due to natural spatial variation in the material properties). In addition, two of the 40 elements, element 5 and 25, are assigned a considerably lower Young's modulus value than the specified nominal value for the beam's material and reflect damages (flaws) in the beam. Figure 3 shows the assumed true Young's modulus values of the 40 elements that make up the finest-resolution finite-element model.

In this exercise, our data consists of the computed nodal displacements (using NIKE3D at resolution 0) plus an added synthetic Gaussian measurement error noise. The nodal displacements are computed at 10 locations along the beam when the following three different sets of forces are applied (as in Glaser et al., 2003):

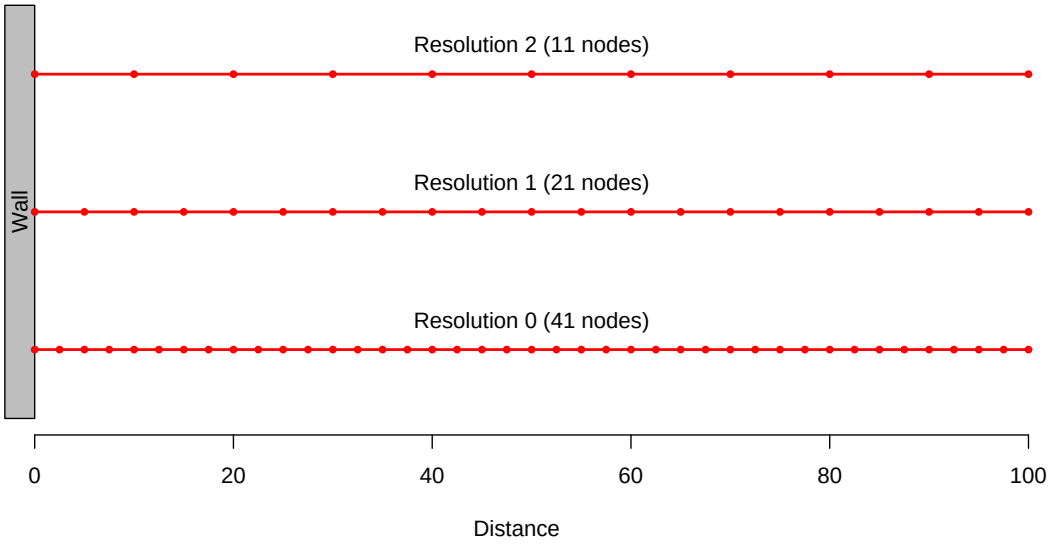

Figure 2: A simple cantilever beam with one end anchored to a wall shown at three resolutions.

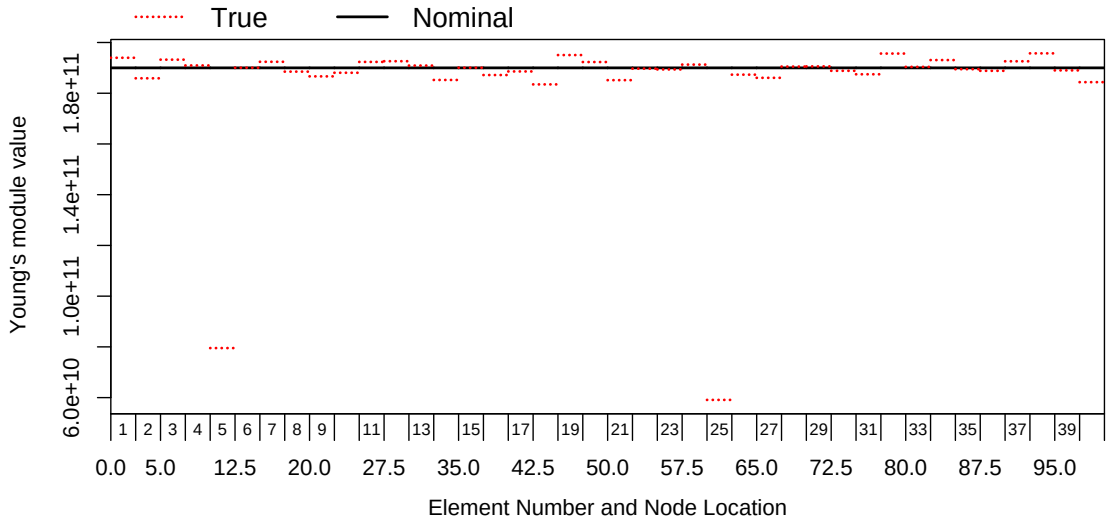

Figure 3: The actual (true) Young's modulus values of the 40 elements that make up the finest-resolution, resolution-0, finite-element model (red dotted line segments). Also shown is the nominal Young's modulus value for the beam's material.

1. A downward force of (relative) strength 1 applied at location 100 (at the free end of beam).

2. A downward force of strength 1/2 applied to locations 50 and 100.

3. A downward force of strength 1/2 applied to location 100 and an upward force of strength 1/2 applied to location 50.

For each of the three sets of forces applied, both vertical (up-down) and angular nodal displacement are measured at locations 10, 20, ..., 100. The observed displacements (the data) are then taken to be equal to the true nodal displacements plus a Gaussian random error with standard deviation of 1.0e-6 in the case of vertical displacement and with standard deviation equal to 1.0e-7 in the case of angular displacement. Figure 4 shows (1) the true nodal displacement, as computed using NIKE3D at resolution 0, (2) the nodal displacement when all elements are assigned a nominal Young's modulus value of 1.9e+11, and (3) the observed displacements. Figure 5 shows then the difference in nodal displacement between a model configured with the true stiffness values and nominal stiffness values, and the difference between the observed displacement and the nominal displacement. To summarize, the current stiffness configuration (with two flaws) results only in a small difference in the computed displacements compared to a model configured with all 40 elements having a nominal, 1.9e+11, Young's modulus value (Figure 4). The task of determining if the characteristics of the beam deviate significantly from a nominal beam and if so, where and how much, becomes even harder with the added measurement noise (Figure 5).

We shall now outline the specification of the prior distribution and the likelihood used at each resolution.

6.1 Specification of the Prior Distribution and the Likelihood

The Prior Distribution

The prior distribution associated with resolution-0 (the finest-resolution) model is fully specified by the parameters of the shifted and scaled Beta distribution for the nominal and the potentially flawed Young's modulus values, along with the binomial rate for the number of potentially flawed elements; see (2) and (3). These parameters are take to be:

$$\begin{aligned} \text{Nominal Beta Param.:} &\quad s^- = 1.80\text{e+}11,\ s^+ = 2.0\text{e+}11,\ \alpha = 5,\ \beta = 5 \\ \text{Flawed Beta Param.:} &\quad s^- = 1.00\text{e+}09,\ s^+ = 1.85\text{e+}11,\ \alpha = 3,\ \beta = 3 \\ \text{Binomial Flaw Rate:} &\quad \lambda = 0.05 \quad (\text{expected } \#\text{ flaws} = 2). \end{aligned} \tag{9}$$

The prior distributions at resolutions 1 and 2 are of the same format as at resolution 0. However, instead of specifying the parameters of the prior distribution for each of the coarser-resolution models independently, they are derived from the prior distribution at resolution 0. This is done with the help of an element aggregation

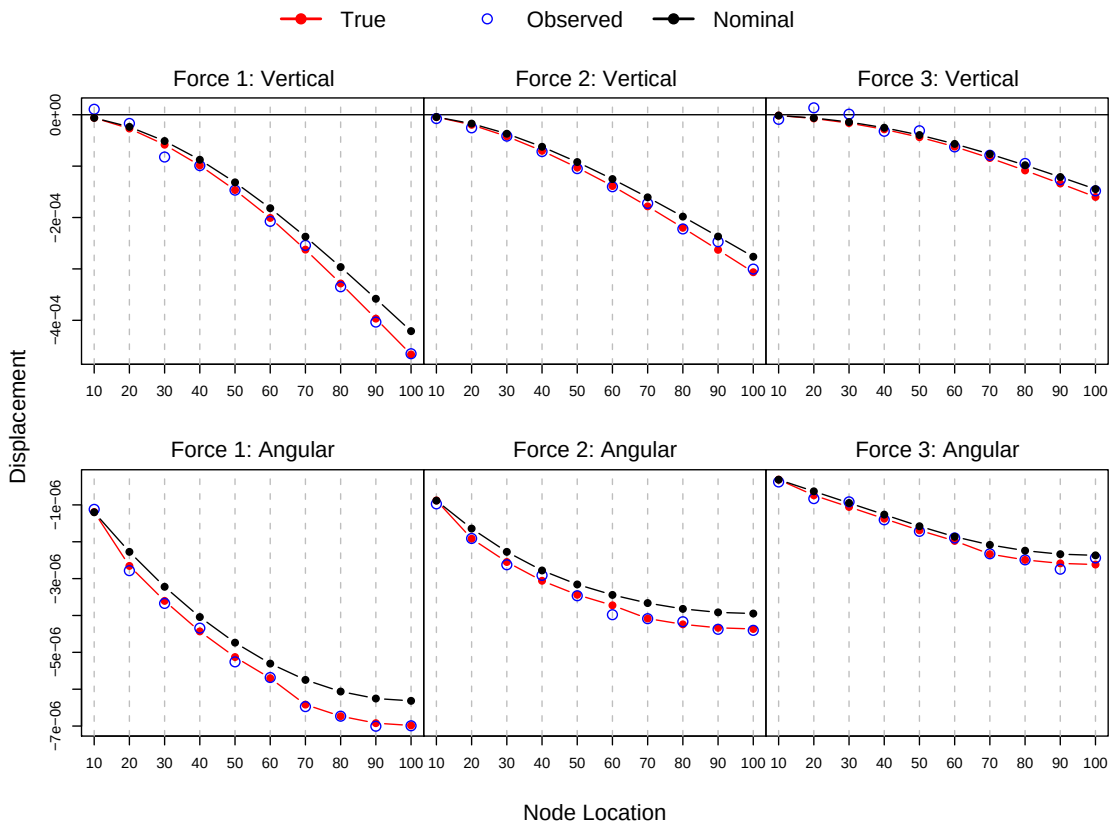

Figure 4: The top three graphs show the vertical true (two flaws), nominal (no flaws), and observed displacement under the three set of forces considered. The bottom row shows the angular displacement.

rule that specifies how a Young's modulus value at a coarse resolution is derived from it's children Young's modulus values at a finer resolution. We consider two aggregation rules; a *weighted arithmetic mean* and a *weighted harmonic mean*, given by:

$$\begin{aligned} s &= w_1 s_1 + w_2 s_1 \quad \text{(weighted arithmetic mean)} \\ s &= \frac{1}{\frac{w_1}{s_1} + \frac{w_2}{s_2}} \quad \text{(weighted harmonic mean)}, \end{aligned} \tag{10}$$

where s_1 and s_2 are the finer-resolution children Young's modulus values and w_1 and w_2 are pre-specified weights. These two aggregation rules correspond to the two suggested between-resolution proposal methods in Section 5.3. In what follows we take $w_1 = w_2 = 1/2$, motivate by equal sized elements. Then, in the case of aggregation via harmonic means, the parameters of the shifted and scaled Beta prior distribution of s are derived to yield an approximation to the true distribution

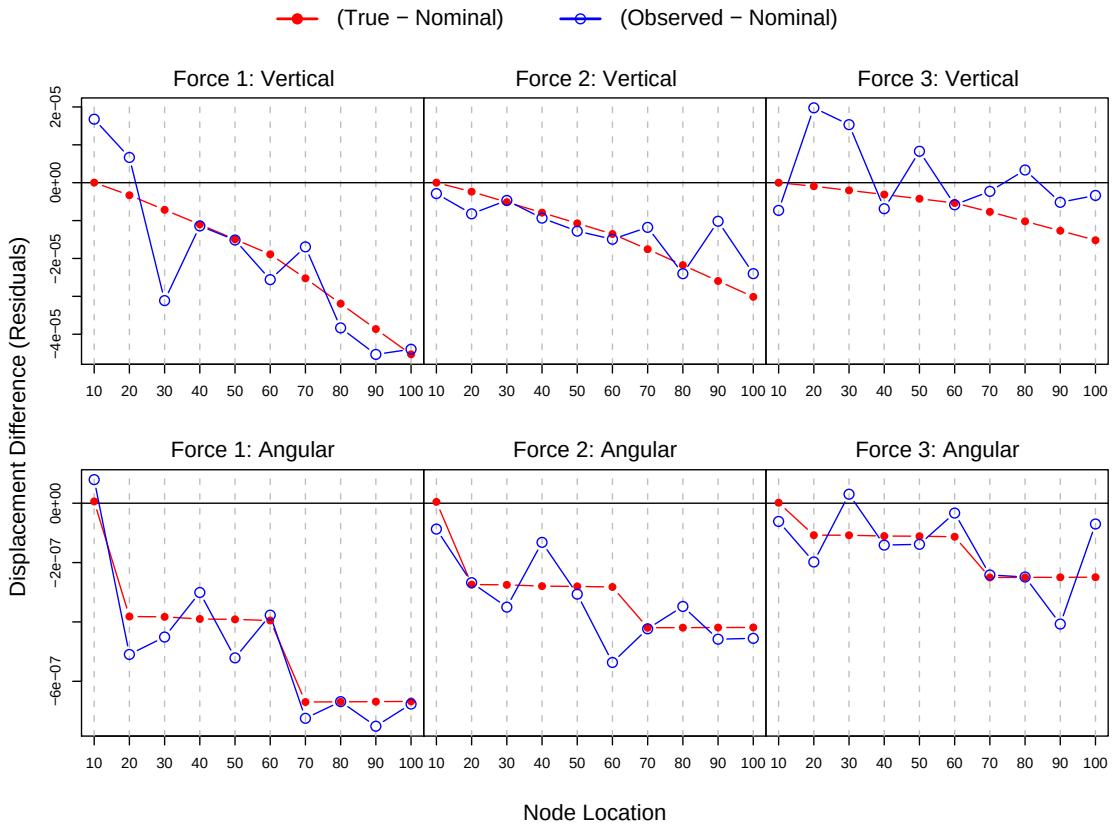

Figure 5: The top three graphs show the difference in vertical displacement between a model configured with two flawed elements (the truth) and one configured with all element as nominal (nominal) and also the difference between the the observed displacement and displacement resulting from the nominal model. The bottom row shows the difference in angular displacement.

of $s = (w_1 s_1^{-1} + w_2 s_2^{-1})^{-1}$ in terms of matching the first two moments (i.e., mean and variance). For example, if both s_1 and s_2 are considered to be nominal, then s is considered to be a nominal with Beta prior approximating the distribution of $(w_1 s_1^{-1} + w_2 s_2^{-1})^{-1}$. Similarly, the Beta prior distribution of a potentially flawed element is derived by approximating the distribution of $(w_1 s_1^{-1} + w_2 s_2^{-1})^{-1}$ when one or both of the finer resolution elements are flawed (as determined by the binomial flaw rate). This yields the following parameters for resolution 1 (derived from resolution-0 priors):

$$\begin{aligned} \text{Nominal Beta Param.:} &\quad s^- = 1.80\text{e+}11,\ s^+ = 2.0\text{e+}11,\ \alpha = 10.5,\ \beta = 10.5 \\ \text{Flawed Beta Param.:} &\quad s^- = 1.00\text{e+}09,\ s^+ = 1.92\text{e+}11,\ \alpha = 4.0,\ \beta = 2.6 \\ \text{Binomial Flaw Rate:} &\quad \lambda = 0.0975 \quad (\text{expected \# flaws} = 1.95). \end{aligned} \tag{11}$$

And for resolution 2 (derived from the resolution-1 priors):

$$\begin{aligned} \text{Nominal Beta Param.:} \quad & s^- = 1.80\text{e+}11,\ s^+ = 2.0\text{e+}11,\ \alpha = 21.5,\ \beta = 20.5 \\ \text{Flawed Beta Param.:} \quad & s^- = 1.00\text{e+}09,\ s^+ = 1.96\text{e+}11,\ \alpha = 5.5,\ \beta = 2.4 \\ \text{Binomial Flaw Rate:} \quad & \lambda = 0.1855 \quad (\text{expected } \#\text{ flaws} = 1.85). \end{aligned} \tag{12}$$

Figure 6 shows the Beta prior distributions at the three resolutions.

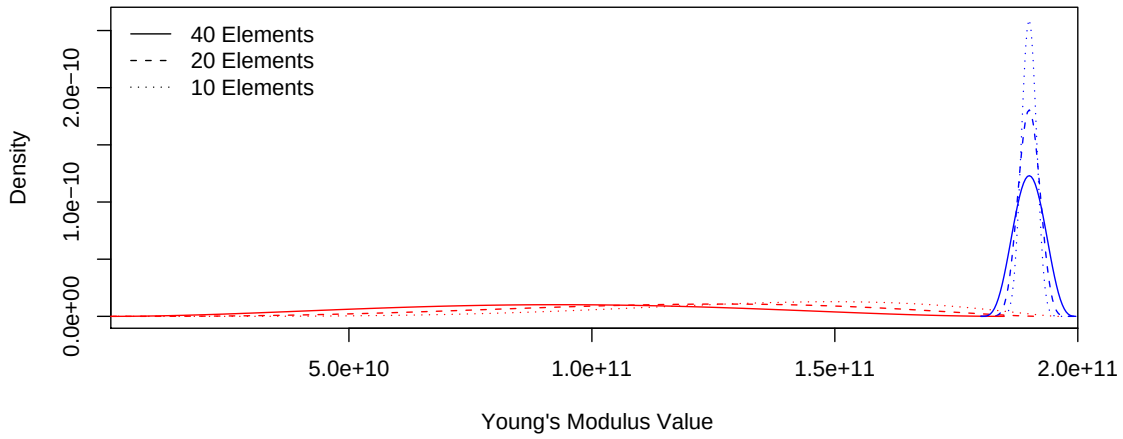

Figure 6: The shifted and scaled Beta prior distributions for the Young's modulus value of a nominal and a potentially flawed element at the three resolutions as derived via simple harmonic mean aggregation rule.

A similar method was employed to derive the parameters for the prior distributions when a simple arithmetic mean aggregation approach was used.

The Likelihood

Given how the data was generated, the likelihood (i.e., the data model in (5)) at each resolution was simply taken as

$$p_r(\mathbf{y} \,|\, \mathbf{s}_r) = \prod_{i=1}^{3} p_{r:i}(\mathbf{y}_i \,|\, \mathbf{s}_r), \quad r = 0, 1, 2,$$

where $\mathbf{y} = \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3\}$ are the observed nodal displacement data under each of the three sets of forces. The force-specific likelihoods are given by,

$$p_{r:i}(\mathbf{y}_i \,|\, \mathbf{s}_r) = \prod_{k=1}^{10} \prod_{j=1}^{2} \varphi(y_{ijk} - \hat{y}_{r:ijk}; \sigma^2_{r:ijk}),$$

where $\mathbf{y}_i = \{y_{ijk} : j = 1, 2,\ k = 1, \ldots, 10\}$, with j indexing the type of nodal displacement (vertical or angular) and k indexing nodal location, $\{\hat{y}_{r:ijk}\}$ are the predicted nodal displacements at resolution r, and $\varphi(\cdot; \sigma^2_{r:ijk})$ is a Gaussian density with mean zero and variance $\sigma^2_{r:ijk}$. The $\{\sigma^2_{r:ijk}\}$ are taken to be given by

$$\sigma^2_{r:ijk} = \psi^2_j + v_r \tau^2_{r:ijk}, \tag{13}$$

where $\{\psi^2_j\}$ are the measurement error variances (ψ_1 = 1.0e-6 and ψ_2 = 1.0e-7), $\{\tau^2_{r:ijk}\}$ are model error variances, and v_r are resolution-specific variance scaling parameters. At the finest resolution we let $\tau^2_{0:ijk} = 0$, reflecting how the data was generated. At resolution 1 and 2, an empirical study was made to investigate the size of the model error with respect to resolution 0. 500 sets of stiffness configurations were generated at the finest resolution from the prior distribution and $\{\hat{y}^{(c)}_{0:ijk}\}$ computed for each configuration $c = 1, \ldots, 500$. Each of the 500 configurations was then aggregated down to resolution 1 and 2 using both arithmetic and harmonic mean, see (10), and the displacements $\{\hat{y}^{(c)}_{r:ijk}\}$, $r = 1, 2$, computed for each aggregation method and stiffness configuration $c = 1, \ldots, 500$. Figure 7 summarizes the result for force configuration 1 ($i = 1$) and the two aggregation methods in terms of the empirical standard deviation

$$\mathrm{sd}_{r:ijk} = \left(\frac{1}{500} \sum_{c=1}^{500} (\hat{y}^{(c)}_{r:ijk} - \hat{y}^{(c)}_{0:ijk})^2 \right)^{1/2}, \quad r = 1, 2.$$

There is a noticeable difference between the arithmetic and the harmonic aggregation approach, with the harmonic approach yielding a considerably better approximation of the fine-resolution model. Using the result of the empirical study, we take $\tau^2_{r:ijk} = \mathrm{sd}^2_{r:ijk}$, $r = 1, 2$.

The argument for the variance scaling parameter v_r in (13) is to have the option of 'increasing' or 'decreasing' the accuracy of the coarser-resolution models. Recall that the coarser-resolution models are mainly used to assist and guide the MCMC sampler at the finest-resolution. As such, 'heating' up the likelihood (i.e., make it flatter) at coarser resolutions might allow the MCMC sampler to move more freely and yield a better mixed and representative sample, as discussed in Section 2.

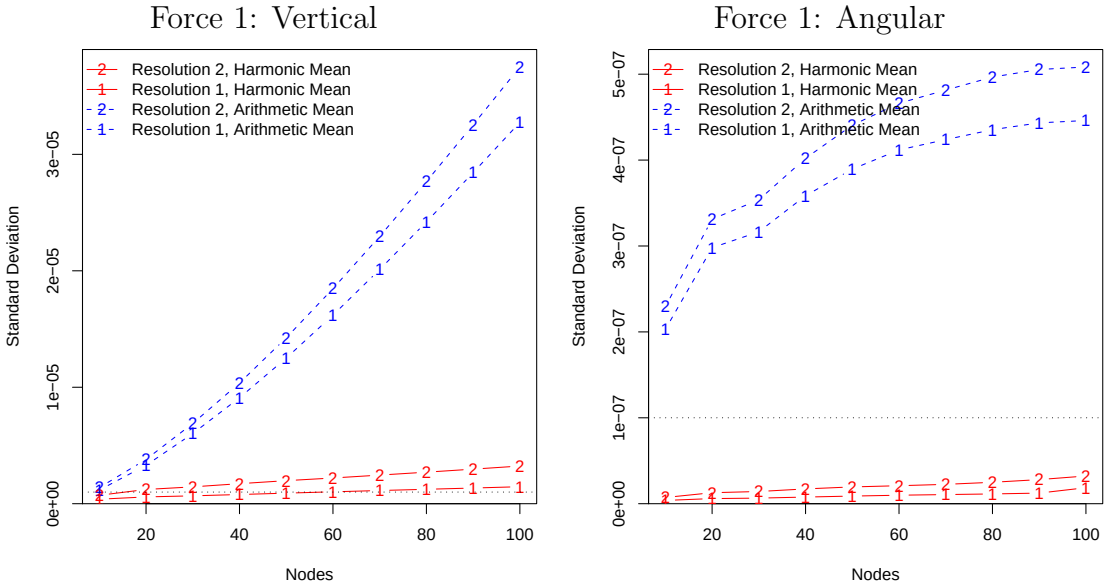

Figure 7: An empirical estimate of the vertical (left graph) and angular (right graph) nodal displacement error computed at resolution 1 and 2, versus at resolution 0. The result of both arithmetic mean and harmonic mean Young's modulus aggregation are shown. Dotted horizontal line shows the measurement error standard deviation used for the two types of displacements.

6.2 Single-Resolution Results

We ran the MCMC sampler for 10,000 iterations at resolution 0, with the stiffness of each element initialized at the nominal value (1.9e+11). Figure 8 shows the configuration of $\{s_i\}$ at iteration 50, 500, 5,000, and 10,000. The sampler "finds" the first flaw (element 5) relative quickly and samples around it, and soon after finds the second flaw (element 25).

Figure 9 shows the sample trace of elements 5 and 25 (both flawed) and of elements 15 and 40 (both nominal). The posterior samples for elements 5 and 25 alter between nominal and flawed states. Element 15, a nominal element located exactly halfway between the two flawed elements, has close to all of is samples in the nominal range, while element 40, a nominal element at the end of the beam, has a larger portion of its samples in the flawed range than element 15. This is somewhat expected as the data provides stronger information about the beam's properties at the beginning of the beam (close to the wall), but are basically non-informative about the last few elements of the beam. Hence, for the last few elements, the sample of the Young's modulus values should reflect the prior distribution. In general, by a crude visual judgment there is a rather short burn-in period, while chain mixing is not seen to be particularly good for nominal elements. This is both due to the samplers design and dimension of the problem (40 elements). That is, by default every fourth proposal is a proposal to resize the Young's modulus value of a nominal element. Hence, for example, approximately every 160 ($= 4\times40$) iteration a proposal is made to resize element 15 (this also hints at how poorly the problem scales with the size of the model, that is, the number of elements). In addition, no attempt was made to tune the proposal distributions in this initial analysis and the acceptance ratio for the four proposal moves is seen to be:

birth/death of a flaw	relocate a flaw	resize a flaw	resize a nominal
0.168	0.607	0.374	0.848

showing a very high acceptance rate for the resizing-a-nominal move, which yields slower mixing among nominal elements. Similarly, the relocate-a-flaw move (with, by default, 90% of the time proposing a relocation to a neighboring location) has a high acceptance rate of 0.607 which can be lowered by lowering the neighboring relocation rate.

Figure 10 shows (top) the sample trace of the total number of (potential) flaws, given by $n^{(j)} = \sum_{i=1}^{40} \ell_i^{(j)}$, where $\ell_i^{(j)}$ is the j-th MCMC sample of ℓ_i, and (bottom) the estimated posterior distribution $\Pr(n \,|\, \text{data})$. The posterior distribution $\Pr(n \,|\, \text{data})$ is estimated by the sample $\{n^{(j)} : j = 1,001, \ldots, 10,000\}$, discarding the first 1,000 samples as a burn-in period (as judged, roughly, by the sample trace of $n^{(j)}$). There is a zero posterior probability of having zero or one flaw, with three flaws having the highest posterior probability. Figure 11 shows estimates of the marginal posterior probabilities of each of the 40 elements being flawed, as given by $\Pr(\ell_i = 1 \,|\, \text{data})$

and estimated by (discarding the first 1,000 samples as burn-in)

$$\frac{1}{9,000}\sum_{j=1,001}^{10,000}\ell_i^{(j)}.$$

Figure 11 also shows the Bayes Factor (BF),

$$\frac{\Pr(\ell_i=1\,|\,\text{data})/\Pr(\ell_i=0\,|\,\text{data})}{\Pr(\ell_i=1)/\Pr(\ell_i=0)};$$

the posterior odds of element i being flawed versus nominal to the prior odds of element i being flawed versus nominal. (A BF between 3 and 12 is consider a positive evidence, a BF between 12 and 150 a strong evidence; Raftery, 1996. It is also helpful to look at $2\log(BF)$, which is on the same scale as the likelihood-ratio in classical statistical testing.) As can be seen, there are positive evidence for flaws at elements 5–8 and at 25–28. Note that the data is observed at nodes 10, 20, ..., 100 (i.e., no data between nodes 10 and 20, and between nodes 60 and 70). A further inspection of the sampler shows that there is typically one flaw at the beginning of the beam that is assigned to one of elements 5–8. For example, the posterior probability of having at least one flaw among the elements 5–8, given by $\Pr(\sum_{i=5}^{8}\ell_i>0\,|\,\text{data})$, is estimated to be 1.0 (i.e., all samples have at least one flaw at elements 5–8), while for locations 25–28 it is estimated to be 0.87. This shows the power of the MCMC approach, where posterior inference can be easily conducted using the generated MCMC sample $\{\mathbf{s}^{(j)},\boldsymbol{\ell}^{(j)}: j=1,001,\ldots,10,000\}$.

Instead of using the samples $\{\boldsymbol{\ell}^{(j)}\}$ to detect potential flaws, one can use the samples $\{\mathbf{s}^{(j)}\}$ directly by looking at the posterior probability of s_i being below a given threshold, say s^o, given by $\Pr(s_i<s^o\,|\,\text{data})$, for any $i=1,\ldots,40$. This probability is simply estimated by counting the number of sampled stiffness values that fall below the given threshold. Figure 12 shows the empirical histogram of the MCMC sample of s_i for elements 4–9 and 24–29, along with a vertical line showing the true Young's modulus value. Figure 12 shows that the estimated size of the flaw at element 5 is closer to the true value than the size of the flaw at element 25. However, in both cases the exact element-location of the flaw is not well isolated within the highly probable range of flawed locations (at elements 5–8 and at elements 25–28; recall Figure 11).

The MCMC sampler was also used to sample at resolution 1 and 2 (with the variance scaling parameter v_r equal to 1.0 in (13) for both resolutions). Figure 13 shows the posterior probability of potentially flawed elements at resolution 1 and 2 (based on 10,000 samples, with the first 1,000 samples discarded). As the elements of the coarser-resolution models are larger, the posterior probability of potential flaws at the respective elements that include the two actual flawed elements increases.

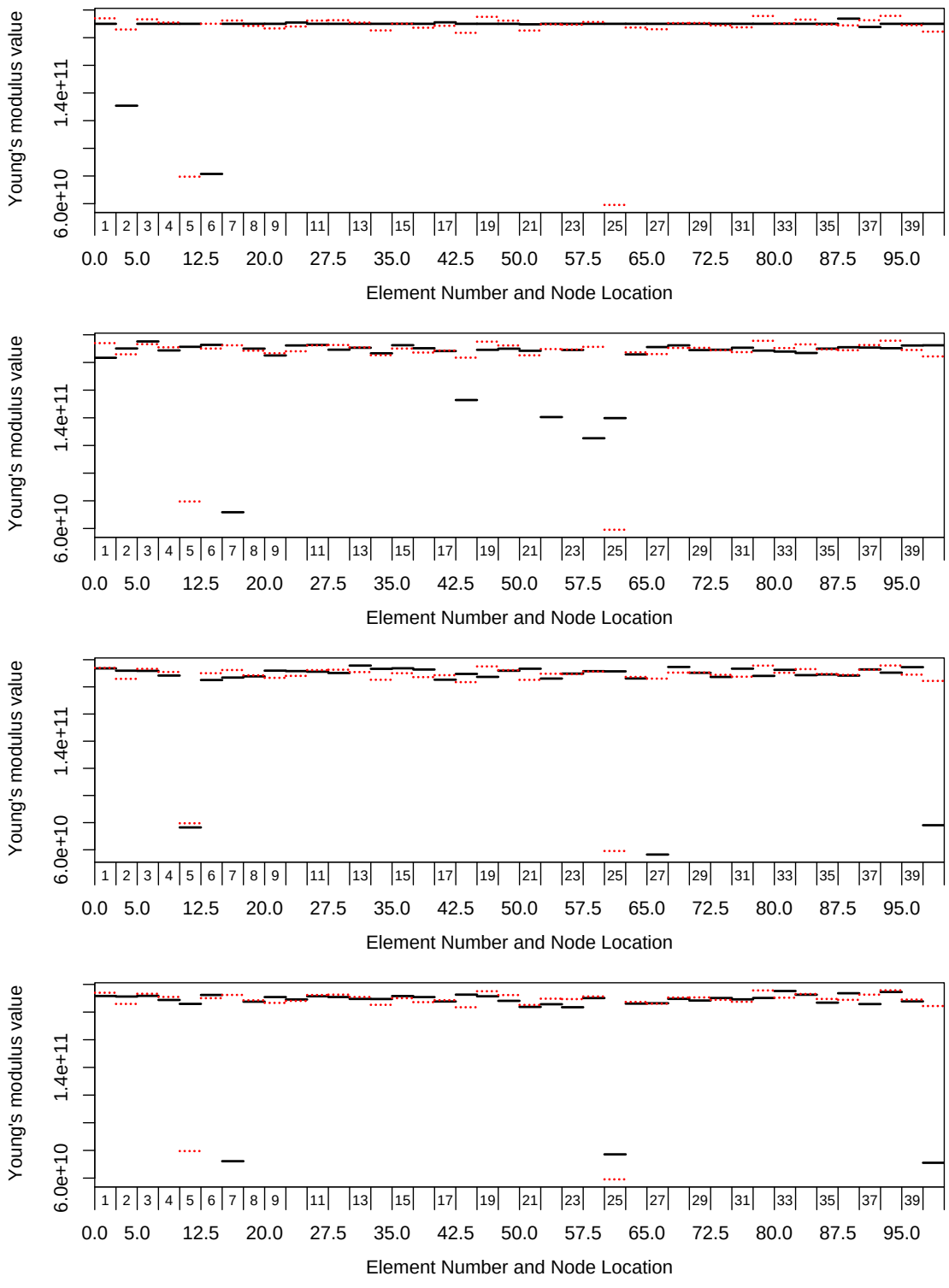

Figure 8: Shown are sampled Young's modulus configurations at iterations (from top to bottom) 50, 500, 5,000, and 10,000 in the case of a single-resolution sampler at resolution 0 (40 elements).

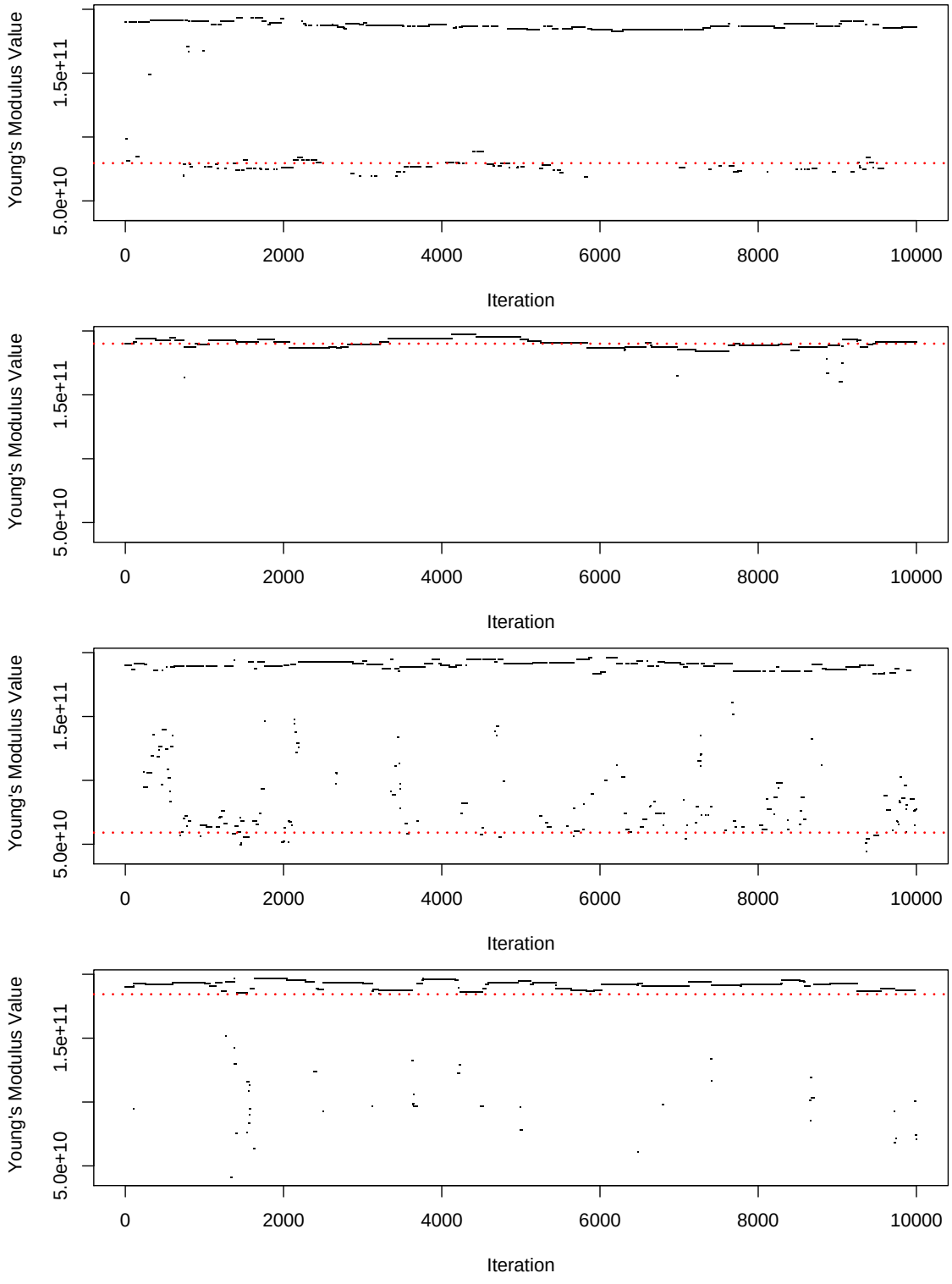

Figure 9: Shown are the sample traces of Young's modulus values for elements (from top to bottom) 5, 15, 25, and 40 in the case of a single-resolution sampler for the 40 element model. Gray dashed line shows the nominal value 1.9e+11 while a red dotted line shows the true Young's modulus value assigned to that particular element.

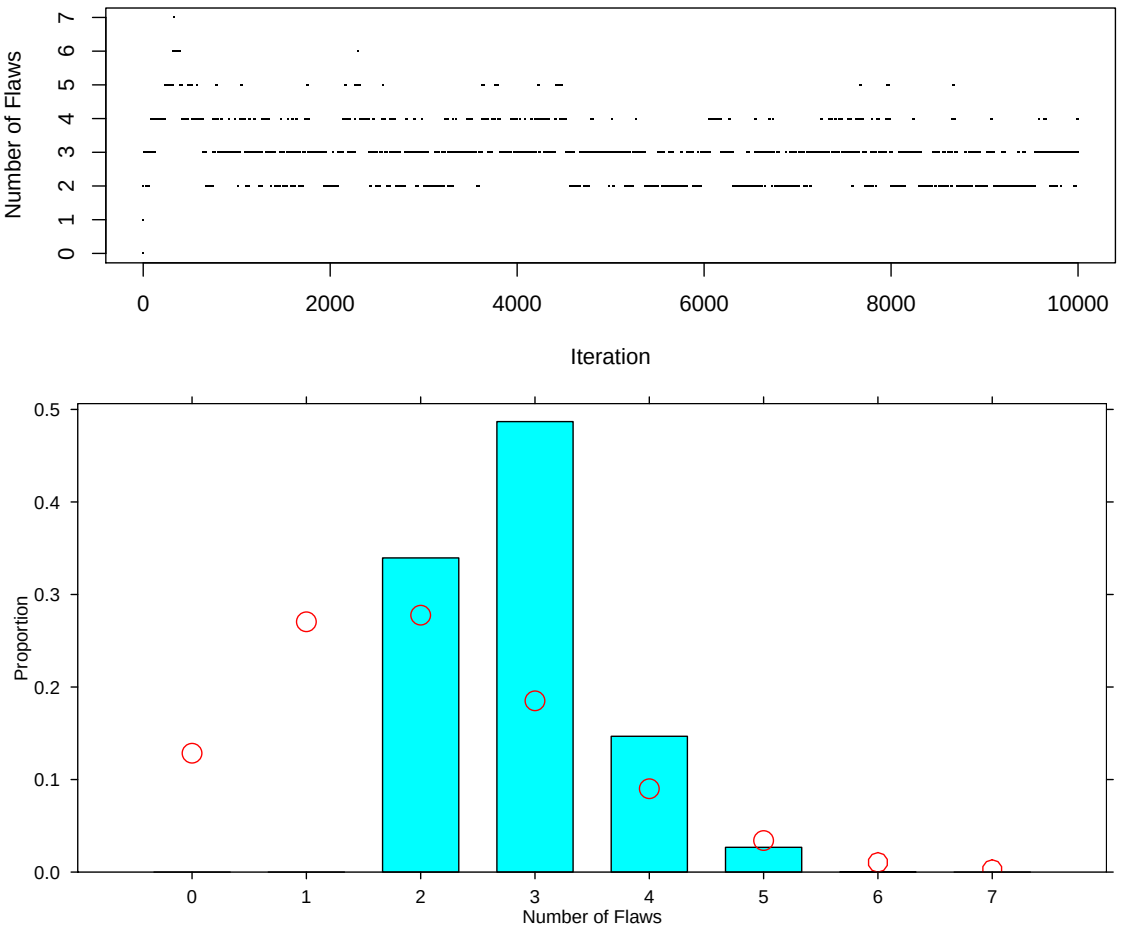

Figure 10: Top, the trace of the total number of (potential) flaws and, bottom, the estimated posterior probability of the number of flaws (bars) along with the prior probability (circles).

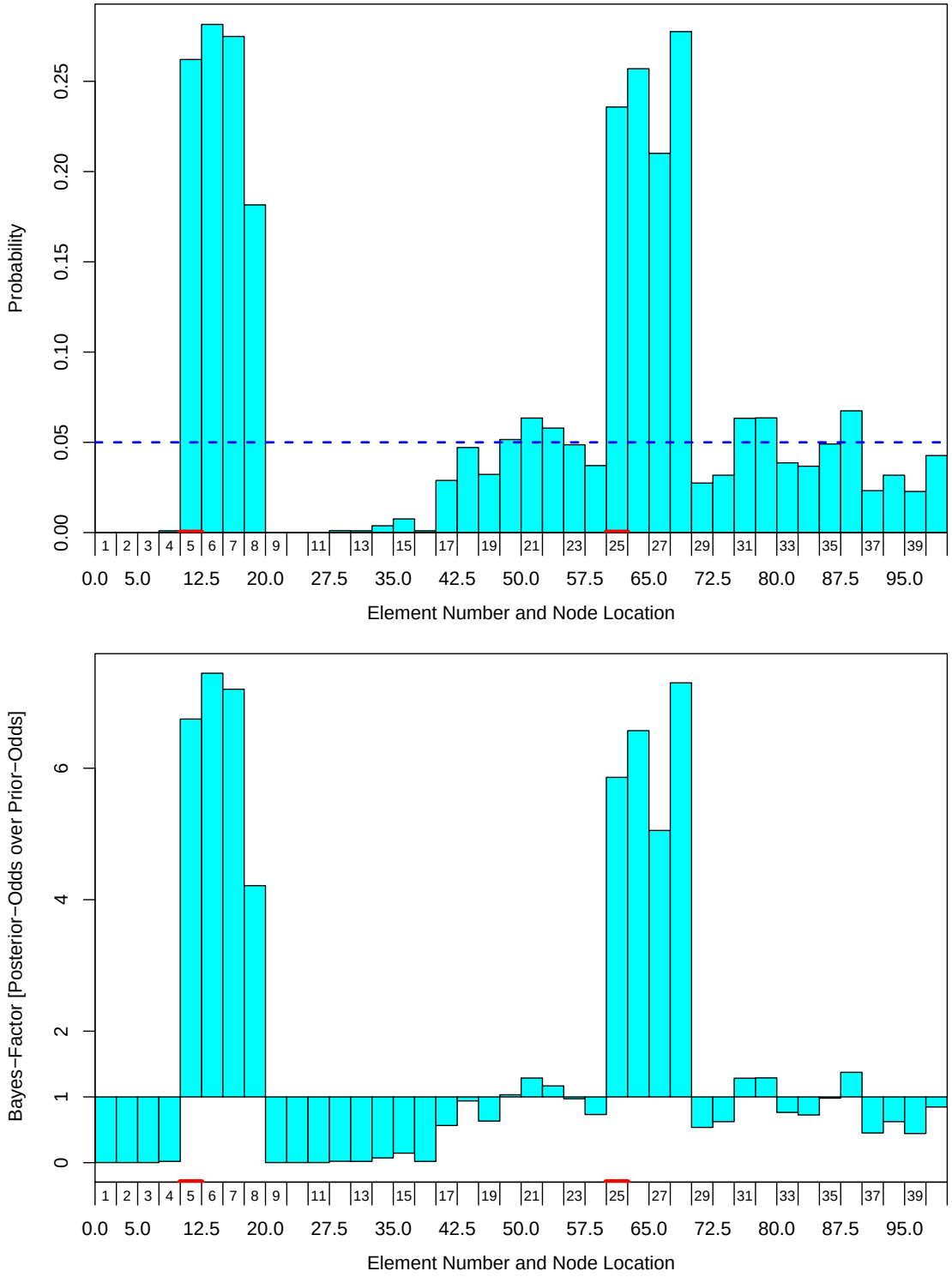

Figure 11: The posterior probability (top) of potentially flawed elements at resolution 0 (the prior probability is shown as a dashed line) and (bottom) the associated Bayes Factor in support of any given element being potentially flawed.

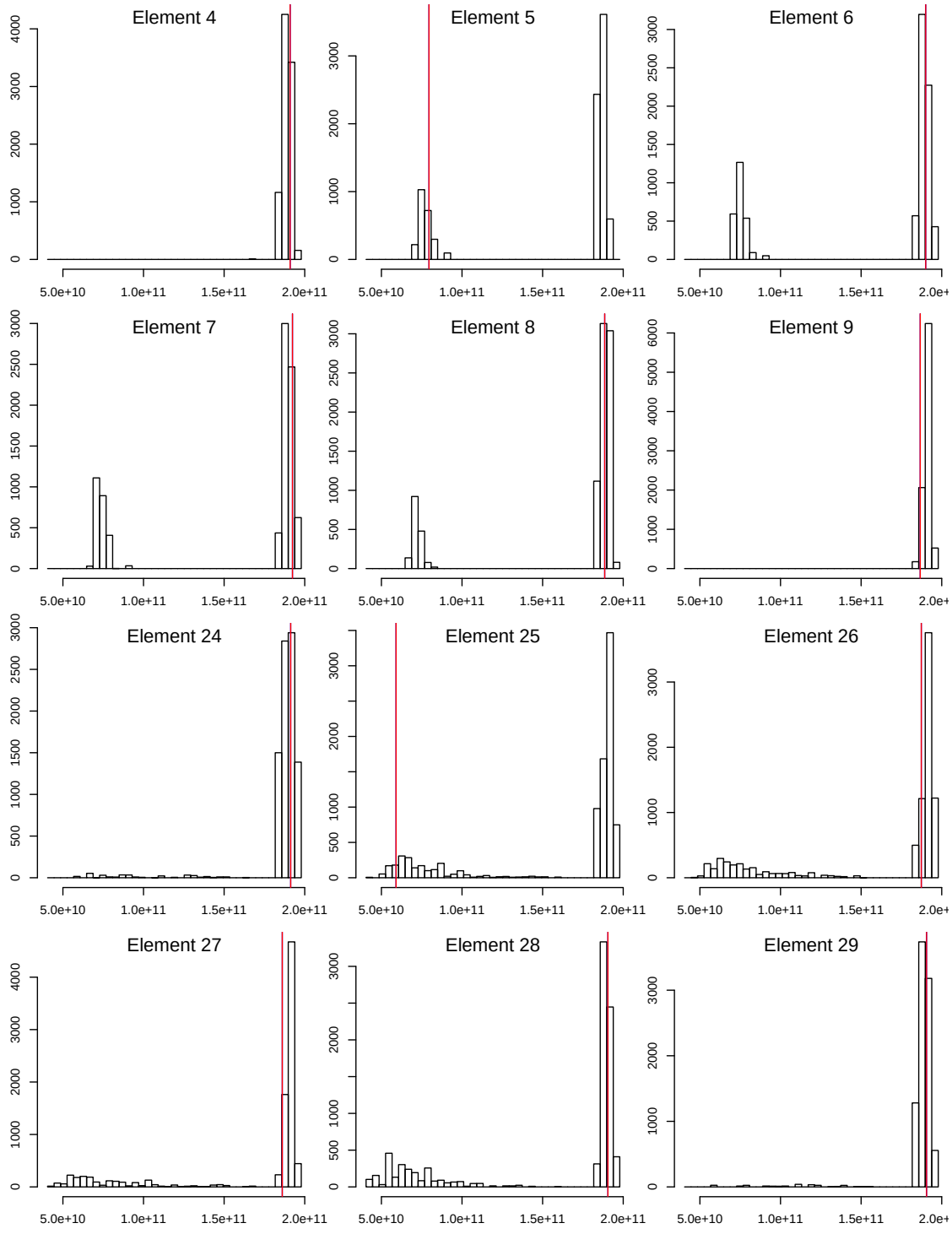

Figure 12: Histograms of MCMC sampled Young's modulus values for elements 4–9 and 24–29 along with a vertical (red) line showing the true Young's modulus value.

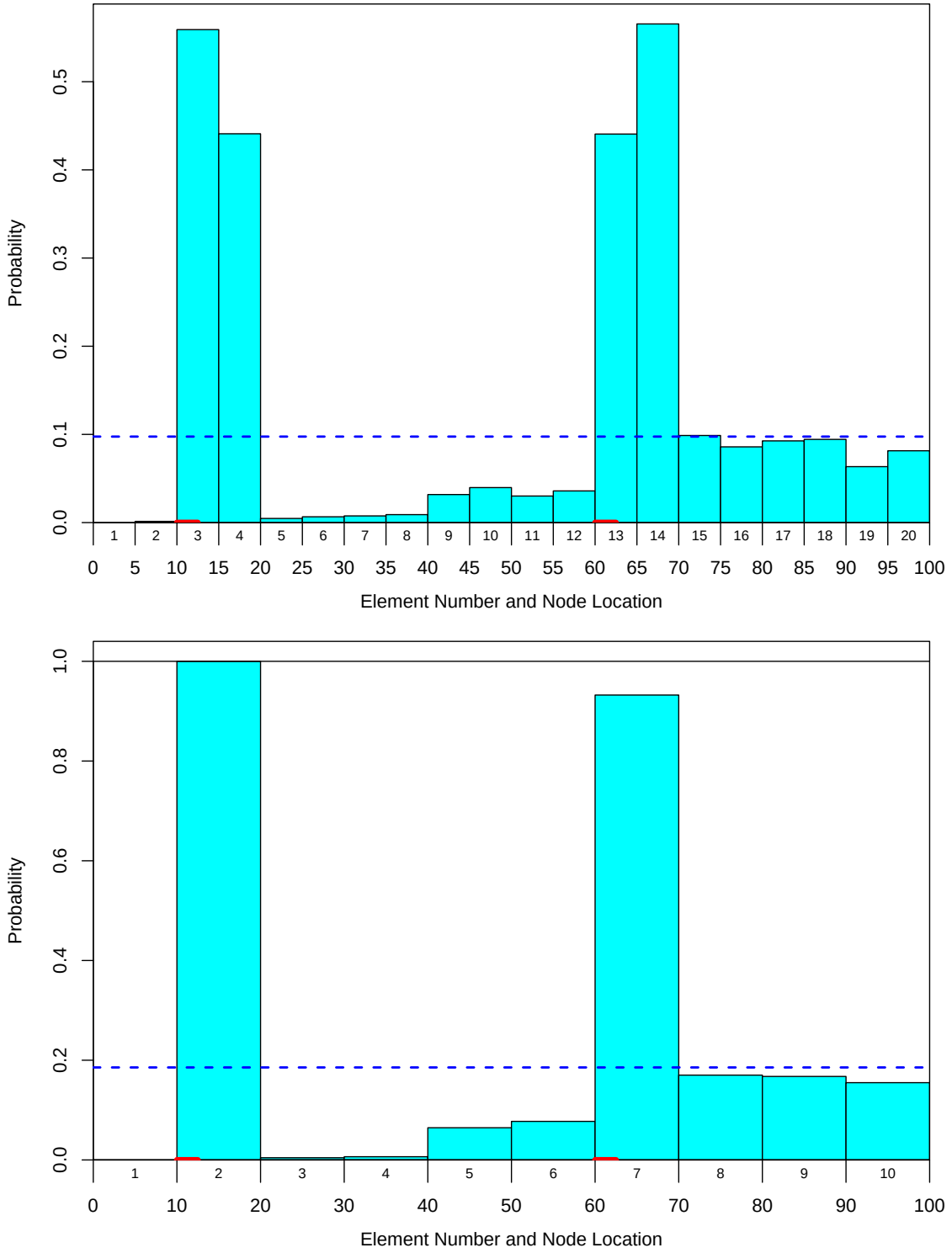

Figure 13: The posterior probability of potentially flawed elements at resolution 1 (top) and at resolution 2 (bottom). Vertical (red) line segments show the location of the two flawed elements at resolution 0.

6.3 Multi-Resolution Results

The multi-resolution MCMC sampler was used to generate 10,000 samples, starting at resolution 2, the coarsest resolution (all with Young's modulus values equal to the nominal value of 1.9e+11). The prior probabilities assigned to each of the three resolutions was (by default) equal. But, once again one can view these prior probabilities as parameters which tune the proportion of the time the sampler stays at each resolution. The between-resolution moves were based on the harmonic-mean aggregation method, as outlined in Section 5.3 with default parameter values. The sampler visited each resolution 3,445, 2,851, and 3,704 times for resolution 0, 1, and 2, respectively. The acceptance rates for going to finer or coarser resolution model were:

	Finer Res	Coarser Res
At Resolution 0	-	0.24
At Resolution 1	0.65	0.79
At Resolution 2	0.30	-

Figure 14 shows how the Markov chain jumps between resolutions. Figure 15 shows four sets of sampled Young's modulus values; samples 38 and 39, when the sampler changes from resolution 2 to 1 for the first time, and samples 90 and 91, when the sampler changes from resolution 1 to 0 for the first time. As can be seen, even as soon as at iteration 90 (when entering resolution 0 for the first time), the Young's modulus sample shows good variation among nominal elements and a flaw configuration that reflects the true flaw locations.

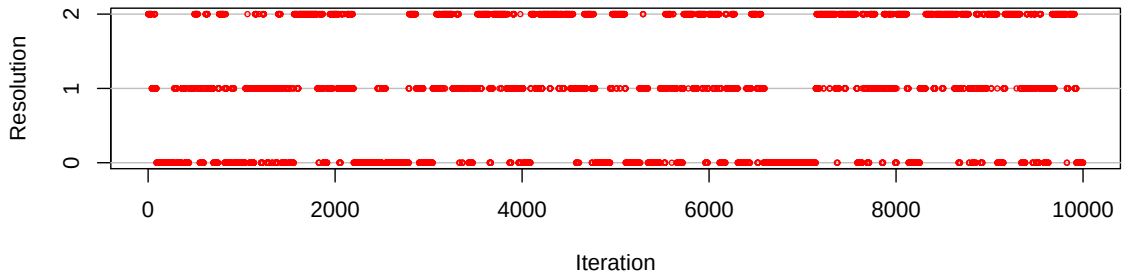

Figure 14: The resolution-trace of the multi-resolution MCMC sampler (resolution 0 is the 40 elements model).

One of the main advantage of the multi-resolution approach over the single-resolution sampler is "quicker" exploration of the parameter space leading to better MCMC mixing. Figure 16 shows the sample traces for the same elements as displayed in Figure 9 for a single-resolution sampler. It is hard to judge from sample traces in Figure 16 if the samples show better mixing than those in Figure 9 (we shall come back to this issue later). Figure 17 and 18 show the posterior evidence of

flawed elements and can be compared to Figure 11 and 13, respectively. Note that a single run (of 10,000 iterations) of the multi-resolution MCMC sampler is used in Figure 17 and 18 while three runs (each of 10,000 iteration) of the single-resolution MCMC sampler are used to draw Figures 11 and 13.

A more formal way to judge mixing, than simply looking at sample traces in Figure 16 is to compute the empirical auto-correlation function (ACF). Figure 19 shows the ACF of the Young's modulus samples from the single-resolution and the multi-resolution sampler for element 5 (a flawed element). Both samplers have a 1,000 iteration burn-in period, but only every other sample is used in the case of the single-resolution sampler. This is done to account for different computational cost by assuming that that evaluation time of the forward model, the NIKE3D finite-element solver, is approximately linear in the number of elements (a very optimistic assumption in favor of the single-resolution sampler). In general we see better mixing in samples from the multi-resolution sampler. This is even more striking for nominal elements, as can be seen in Figure 20, which shows the ACF for element 15 (a nominal element).

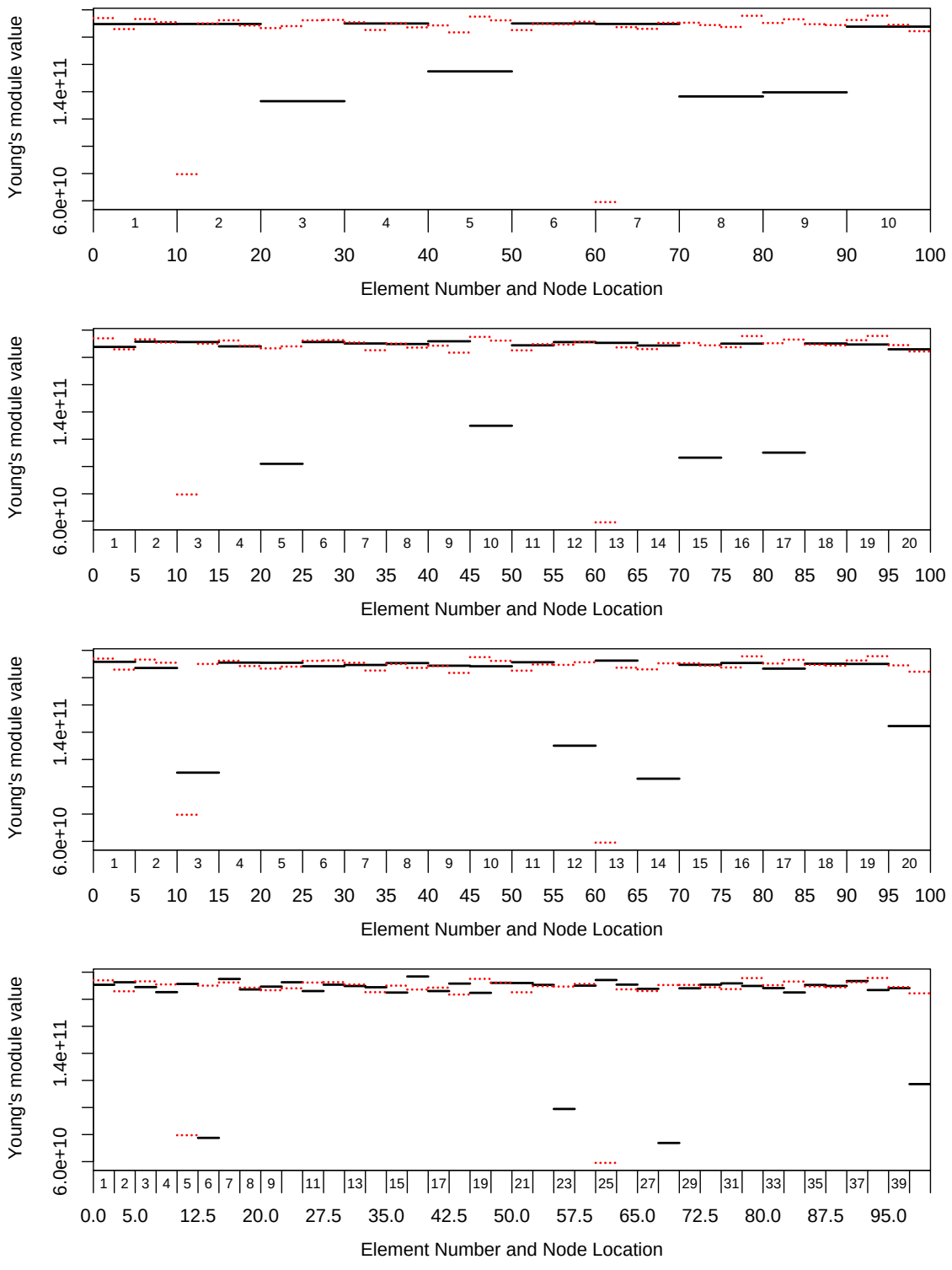

Figure 15: Shown are sampled Young's modulus configuration at iterations (from top to bottom) 38, 39, 89, and 90 in the case of a multi-resolution sampler (the sampler starts at resolution 2 and enters resolution 1 for the first time at iteration 39 and resolution 0 for the first time at iteration 90).

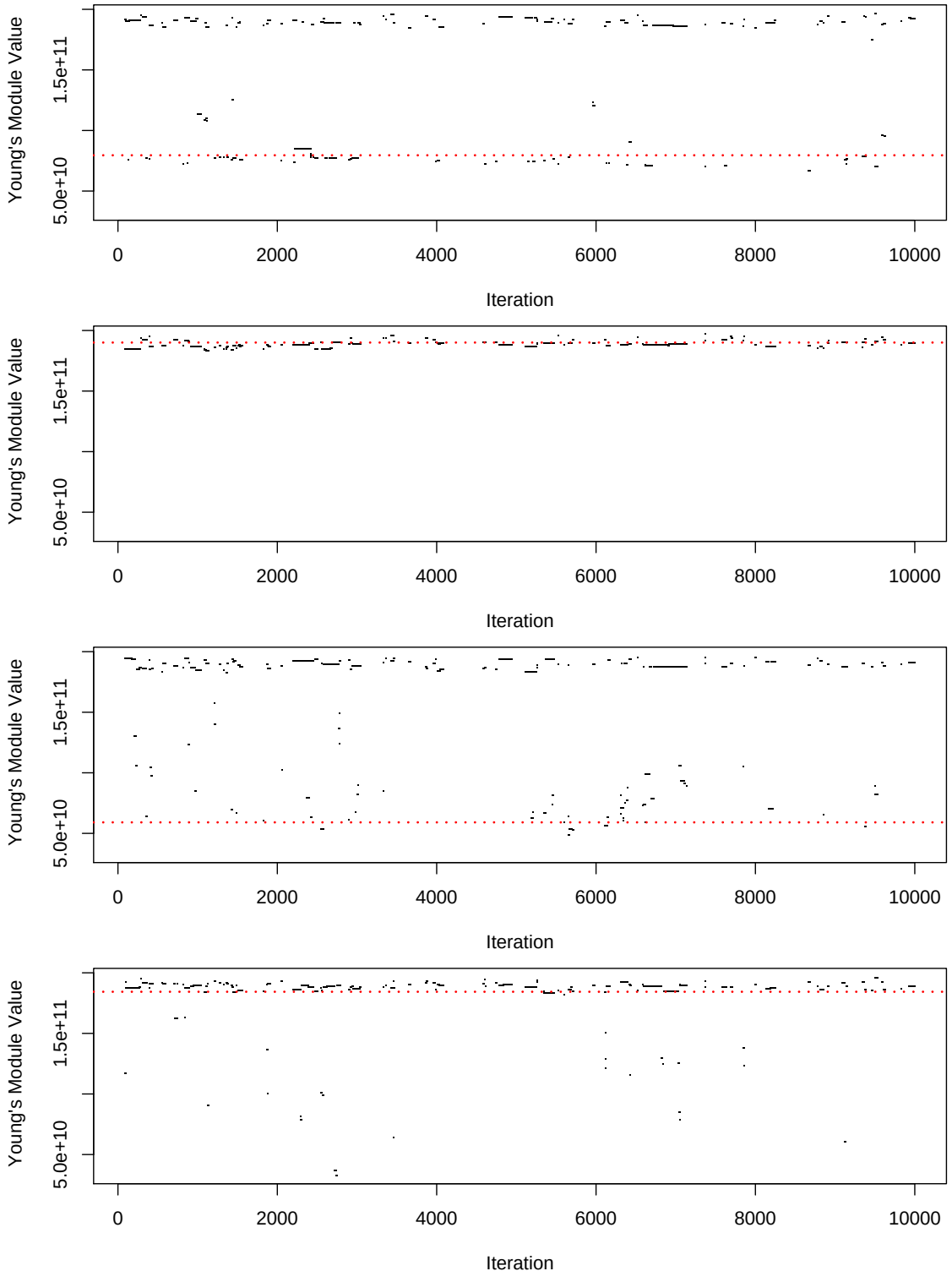

Figure 16: Shown are the sample traces of Young's modulus values for elements (from top to bottom) 5, 15, 25, and 40 (at resolution 0) generated by the multi-resolution sampler. Gray dashed line shows the nominal value 1.9e+11 while a red dotted line shows the true Young's modulus value assigned to that particular element.

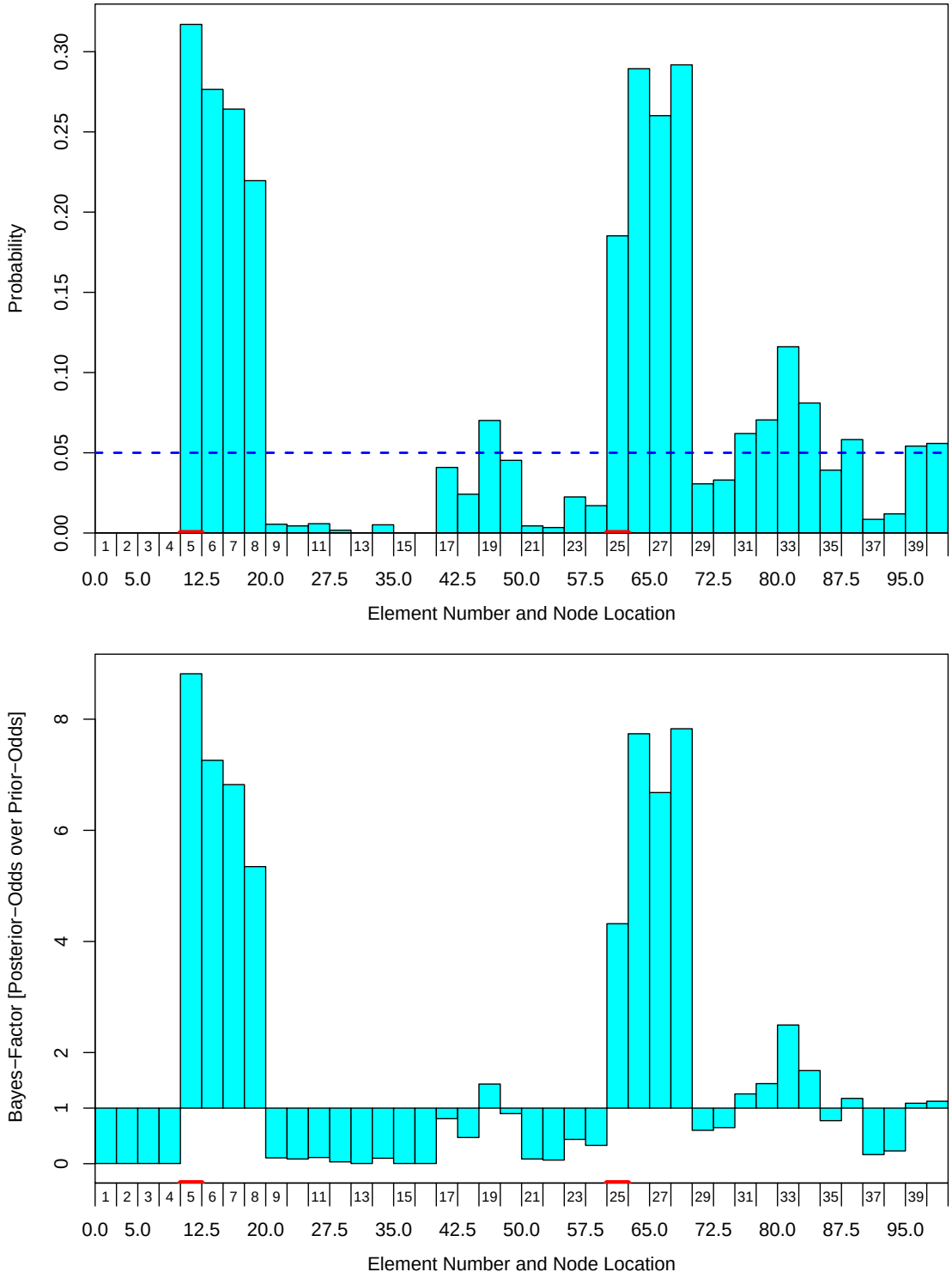

Figure 17: The posterior probability (top) of potentially flawed elements at resolution 0 using the multi-resolution MCMC sampler (the prior probability is shown as a dashed line) and (bottom) the associated Bayes Factor in support of any given element being potentially flawed.

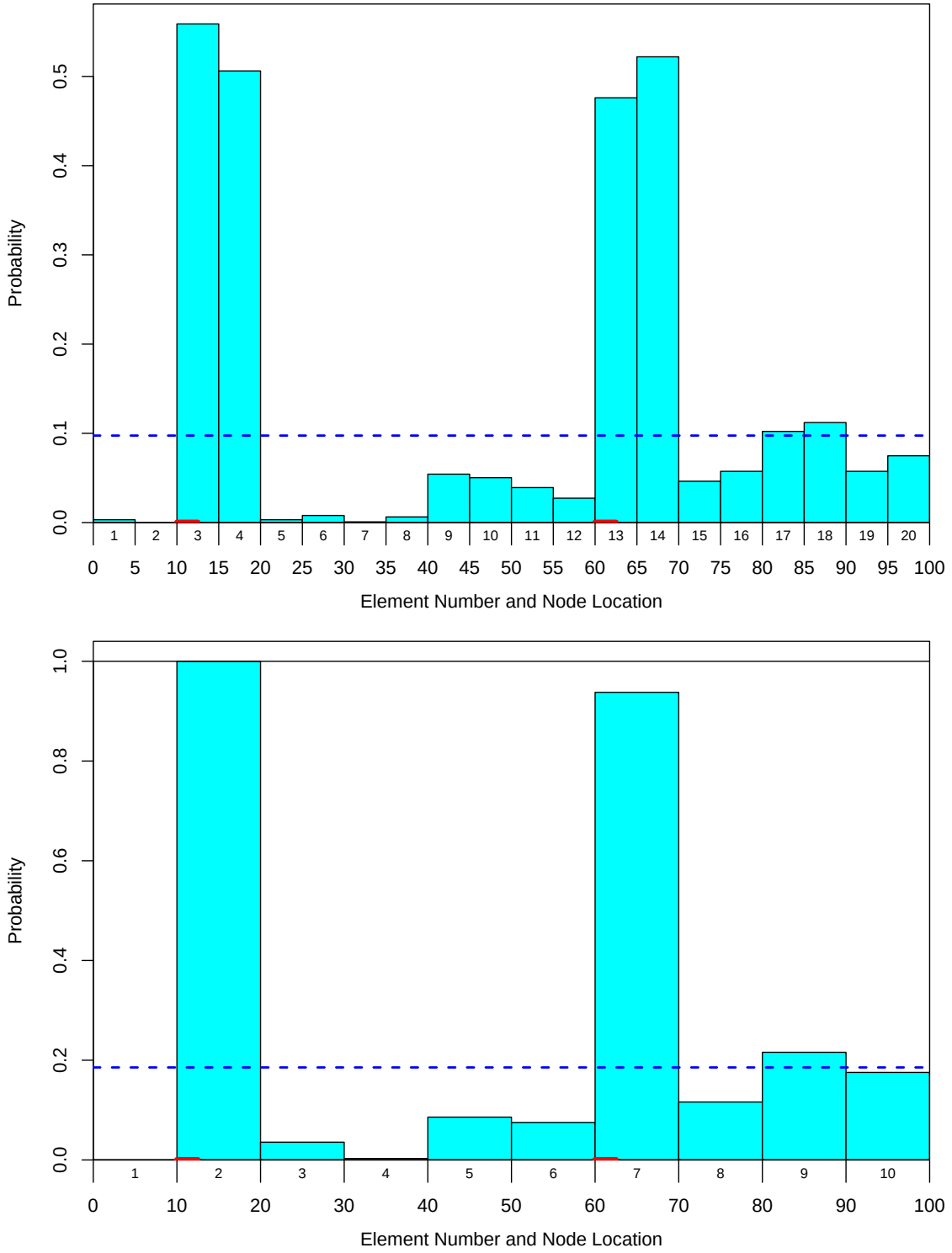

Figure 18: The posterior probability of potentially flawed elements at resolution 1 (top) and at resolution 2 (bottom) using the multi-resolution MCMC sampler. Vertical (red) line segments show the location of the two flawed elements at resolution 0.

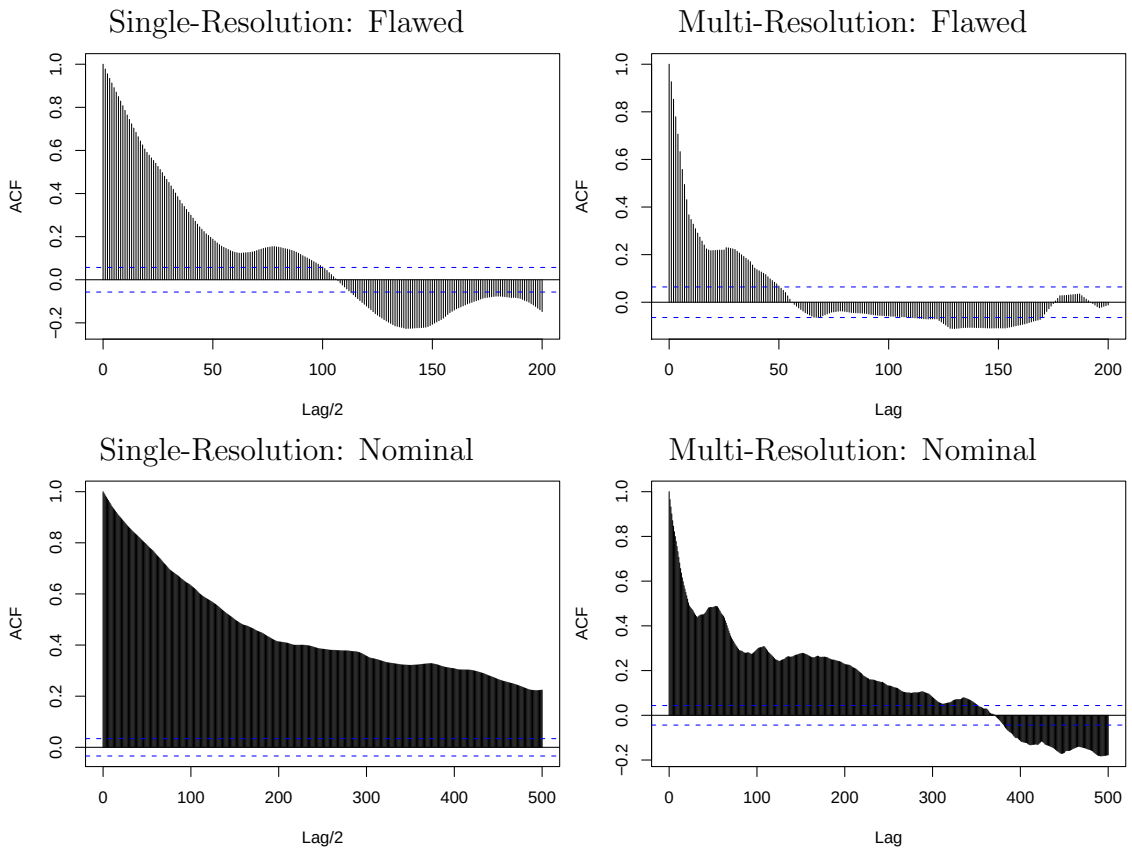

Figure 19: Auto-correlation in Young's modulus samples for element number 5 at resolution 0 (a flawed element) generated by the single-resolution MCMC sampler and the multi-resolution MCMC sampler. The top row shows auto-correlation in $\{s_5(j) : \ell_5^{(j)} = 1\}$ and the bottom row shows the auto-correlation in $\{s_5(j) : \ell_5^{(j)} = 0\}$.

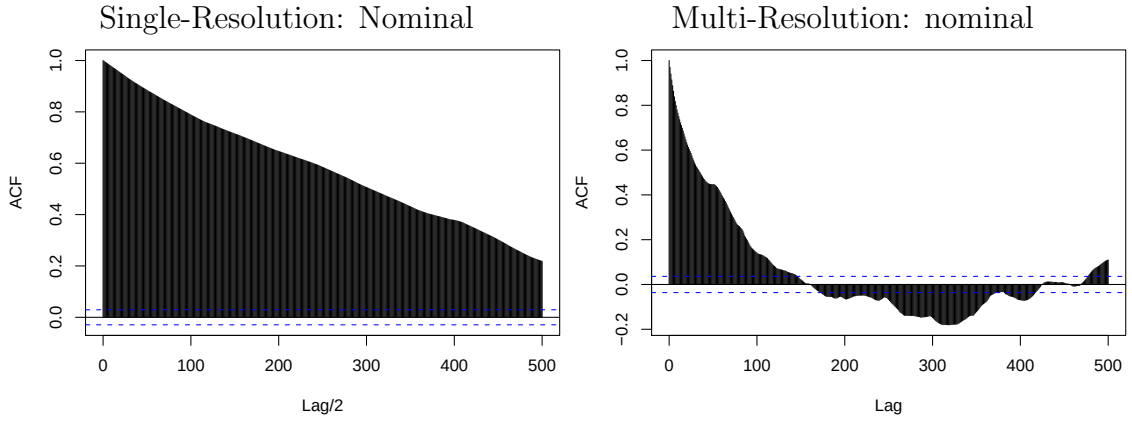

Figure 20: Similar to Figure 19, except in the case of element 15 at resolution 0.

7 Discussion

This initial attempt of using a multi-resolution MCMC inverse approach to perform flaw detection in a physical structure shows promise. The main goal of the multi-resolution methodology is to improve on mixing and reduce the computational cost associated with the forward model (the NIKE3D solver in this case). Improved mixing allows one to reduce the MCMC sample size needed for meaningful system inference, hence reducing the overall computational cost. Such improvement is accomplished by alternating between small- and large-scale exploration of the parameter space. It is possible to conduct both small- and large-scale exploration at a single resolution by moving "blocks" of parameters in a single step. However, we can reduce the computational cost even further by conducting the large-scale exploration at coarser resolutions with a speedier forward model. As a byproduct, inference is performed simultaneously at multiple resolutions.

Our application to the 1-dimensional cantilever beam of Glaser et al. (2003) is in many ways too "small" to demonstrate the potential benefits of the multi-resolution approach. Not only is the dynamic of the beam well-described by a relatively coarse model (i.e., 10 elements), but additionally the coarsening of the 1-dimensional 40 element model results only in modest dimension reduction (from 41 to 21 nodes, and from 21 to 11 nodes — approximately 1/2 reduction in each step). For larger 3-dimensional structures, coarsening results in a greater reduction in the number of nodes. For example, one gains approximately $(1/2)^3 = 1/8$ reduction in total number of nodes using a similar scheme to that used for the 1-dimensional cantilever beam.

To end on a little example, assume that it is only computationally feasible, due to time restriction, to evaluate a fine-resolution 3-dimensional finite-element model 1,000 times. A single-resolution MCMC analysis limited to 1,000 iteration may not even pass the burn-in period, yielding it useless. Now, assume that the forward model scales linear in the number of nodes (a very mild assumtion). As such, a single fine-resolution model evaluation can be replaced with 8 medium-resolution model evaluations (assuming 1/8 reduction in the total number of nodes) or with 64 coarse-resolution model evaluations. Hence, the 1,000 fine-resolution runs can be replaced with, for example, 700 fine-resolution runs, 1,600 medium-resolution runs, and 6,400 coarse-resolution runs. That is, using a multi-resolution MCMC approach would allow one to pass the burn-in period, using coarse- and medium-resolution models, and yield a well mixed sample for the fine-resolution model.

References

Aines, R. L., et al. (2002). The Stochastic Engine initiative: improving prediction of behavior we in geologic environment we cannot directly observe. Technical Report UCRL-ID-148221, Lawrence Livermore National Laboratory, Livermore.

Andrieu, C., Djurieé, P., & Doucent, A. (2001). Model selection by MCMC computation. *Signal Processing*, *81*, 19–37.

Andrieu, C., Doucent, A., fitzgerald, W. J., & Pérez, J.-M. (2000). Bayesian computational approahces to model selection. In W. J. Fitsgerald, R. L. Smith, A. T. Waldon, & P. C. Young (Eds.), *Nonlinear and Nonstationary Signal Processing*. Cambridge: Cambridge Press.

Bernardo, J. M. & Smith, A. F. M. (1994). *Bayesian Theory*. Wiley.

Besag, J. (1997). Discussion of "Bayesian analysis of maixtures with an unknown number of components", by S. Richardson and P. Green. *Journal of the Royal Statistical Society, Series B*, *59*, 774.

Besag, J. (2001). Markov chin Monte Carlo for statistical inference. Technical Report No. 9, Center for Statistical and the Social Sciences, University of Washington.

Brooks, S. P., Guidici, P., & Roberts, G. O. (2003). Efficient construction of reversible jump Markov chain Monte carlo proposal distributions (with discussion). *Journal of the Royal Statistical Society, Series B*, *65*, 3–55.

Charlin, B. & Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo. *Journal of the Royal Statistical Society, Series B*, *57*, 473–484.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. E. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall.

Glaser, R. E., Hanley, W. G., Lee, C. L., & Nitao, J. J. (2003). A Markov chain Mnote Carlo based method for system identification. Technical Report UCRL-ID-150494, Lawrence Livermore National Laboratory.

Godsill, S. (2001). on the relationship between MCMC model uncertainty methods. *Journal of Computational and Graphical Statistics*, *10*, 230–248.

Green, P. (2002). Trans-dimensional Markov chain Monte Carlo. In *Highly Structured Stochastic Systems*. Oxford: Oxford University Press. To be publised.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*, 711–732.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. New York: Springer.

Newmark, R. L., et al. (2002). Stochastic Engine: direct incorporation of measurements into predictive simulations. In *Proceeding of the International Groundwater Symposium "Bridging the gap between measurement and modeling in heterogeneous media"*, number UCRL-JC 145116, (pp. 25–28)., Berkeley, CA.

Raftery, A. E. (1996). Hypothesis testing and model selection. In W. R. Gilks, S. Richardson, & D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice* (pp. 165–188). London: Chapman & Hall.

Richardson, S. & Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society B*, *59*, 731–792.

Waagepetersen, R. & Sorensen, D. (2001). A tutorial on reversible jump MCMC with a view toward applications in QTL-mapping. *Inernational Statistical Review*, *69*, 49–61.