A methodology for sensitivity analysis in inverse problems—application to a palaeoclimate study

Sourabh Bhattacharya*,†,‡

Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, India

SUMMARY

We present a methodology for implementing sensitivity analysis in the case of inverse problems. This is motivated by a problem in palaeoclimatology, where interest lies in reconstruction of past climates from fossil data. In such situations fossil data are modelled as functions of climate variables, since in general ecological terms, variations in climate drive variations in fossil. However, prediction of climate variables are of interest, indicating the inverse nature of the problem.

Technically, given a data set $\{x_i, y_i\}; i = 1, \ldots, n$, and a probability distribution $p(y_i \mid x_i, \theta)$, where θ is the set of model parameters, the problem is 'inverse' if, given a new observed y, prediction of the corresponding unknown x is required. On the other hand, the more usual forward problem considers prediction of y for given x. Sensitivity analysis in forward problems is about examining if the results are robust with respect to changes of the prior on θ. This has received a lot of attention in the statistical literature. However, in the case of inverse problems, a prior on x is required, in addition to the prior on θ. This complicates sensitivity analysis, at least operationally, and, to our knowledge, there exists no work in the literature that addresses this. In this paper, we propose a new methodology for sensitivity analysis in inverse problems, and discuss its application to a real, complex, palaeoclimate problem.

KEY WORDS: climate variable; importance weights; Poisson regression; Reweigh; saturated posterior; temporal smoothness

1. INTRODUCTION

In serious Bayesian analysis it is desirable that the conclusions of the analysis are not heavily dependent on an arbitrary choice of the prior distribution of model parameters. To check whether this is the case, it is necessary to obtain the posterior results under all possible reasonable priors. Since in complex realistic Bayesian analyses, posterior analysis typically proceeds via highly computer intensive Markov chain Monte Carlo (MCMC) (see, for e.g. Tierney (1994)) methods, obtaining posteriors under all possible reasonable prior specifications is computationally burdensome. In the case of forward problems, where given data $(X, Y) = \{x_i, y_i\}; i = 1, \ldots, n$ and probability density $p(y_i \mid x_i, \theta)$ (θ being the set of model parameters), prediction of y for a given x is of interest, methods based on importance sampling has been proposed to ease computational burden. In this paper, we show that such methods are not applicable to inverse problems, where, instead of prediction of y for given x, interest lies in prediction of x for given y. To our knowledge, there exists no work that addresses sensitivity analysis in inverse problems; however, there is some limited discussion on this in Haslett *et al.* (2006).

We develop an approach which is a variant of Importance Resampling MCMC (IRMCMC), originally introduced by Bhattacharya and Haslett (2007) to solve the cross-validation problem in inverse settings. The methodology combines Importance Resampling (IR) (see, e.g. Rubin (1988)) with MCMC to radically reduce high dimensionality. We demonstrate that our approach is computationally cheap and is highly suitable for exploring multimodal posteriors. We further demonstrate that it is highly amenable to parallel implementation. All these have important consequences, particularly in the case of palaeoclimate reconstruction problems. In brief, palaeoclimatology involves reconstruction of past climates (X) from fossils (Y) in lake sediment. The model in this situation posits that variations in organisms (fossil data) are implied by variations in past climates. So, Y is modelled as dependent on X. However, since interest lies in prediction of past climates, given fossil data, one must predict X from Y. For recent works on inverse modelling in the case of palaeoclimate studies, see Vasko *et al.* (2000), Haslett *et al.* (2006), Bhattacharya (2006). Apart from palaeoclimate problems, of course, there are many more examples of inverse problems in environmental science and physics. For some recent works relating to inverse problems in science other than palaeoclimatology, see, for example Haario *et al.* (2004), Cornford *et al.* (2004). These clearly demonstrate the importance of inverse problems in science, and hence the importance of sensitivity analysis in inverse problems.

Before proceeding with sensitivity analysis, we first review the differences between forward and inverse problems, and indicate why sensitivity analysis in the latter case requires careful attention.

1.1. Forward problems

Assuming conditional independence, the likelihood of Y given X, θ is given by

$$L(Y, X, \theta) = \prod_{j=1}^{n} p(y_j \mid x_j, \theta) \tag{1}$$

The posterior distribution of θ given (X, Y) is given by

$$\begin{aligned} \pi(\theta \mid X, Y) &\propto \pi(\theta) L(Y, X, \theta) \\ &= \pi(\theta) \prod_{j=1}^{n} p(y_j \mid x_j, \theta) \end{aligned} \tag{2}$$

Suppose that, prediction is needed for the unobserved y, where x has been observed. Denoting unobserved y by $\tilde{y}$, we note that it is required to compute a posterior predictive distribution of $\tilde{y}$. This is given by

$$\pi(\tilde{y} \mid X, Y, x) = \int p(\tilde{y} \mid x, \theta) \pi(\theta \mid X, Y) \mathrm{d}\theta$$

$$\propto \int \pi(\theta)p(\tilde{y} \mid x, \theta) \prod_{j=1}^{n} p(y_j \mid x_j, \theta)\mathrm{d}\theta \tag{3}$$

As an example, let us consider a Poisson regression problem, where, for $i = 1, \ldots, n$, $y_i \sim \text{Poisson}(\theta x_i)$. Using a flat prior on θ it follows from Equation (3) that

$$\pi(\tilde{y} \mid X, Y, x) \propto \frac{x^{\tilde{y}}}{\tilde{y}!} \frac{\Gamma\left(\tilde{y} + \sum_{j=1}^{n} y_j + 1\right)}{\left(x + \sum_{j=1}^{n} x_j\right)^{\left(\tilde{y}+\sum_{j=1}^{n} y_j+1\right)}}$$

This is the well-known forward problem. Note that, only prior on model parameter θ is needed in this case.

1.2. Inverse problems

On the other hand, if it is of interest to predict x (which is unobserved, and denoted by $\tilde{x}$), given observed y, then the problem is inverse. In this case, it is necessary to compute the posterior predictive distribution of $\tilde{x}$, which is given by

$$\begin{aligned}\pi(\tilde{x} \mid X, Y, y) &= \int \pi(\tilde{x} \mid y, \theta)\pi(\theta \mid X, Y, y)\mathrm{d}\theta \\ &\propto \int \pi(\theta)\pi(\tilde{x} \mid \theta)p(y \mid \tilde{x}, \theta) \prod_{j=1}^{n} p(y_j \mid x_j, \theta)\mathrm{d}\theta \end{aligned} \tag{4}$$

It is to be noted that, unlike in the case of forward problems, the posterior predictive (Eq. 4) requires a *prior* distribution for $\tilde{x}$. In Equation (4) this is indicated by $\pi(\tilde{x} \mid \theta)$. For the Poisson regression problem, using flat priors on both $\tilde{x}$ and θ it follows that

$$\pi(\tilde{x} \mid X, Y, y) \propto \frac{\tilde{x}^{y}}{\left(\tilde{x} + \sum_{j=1}^{n} x_j\right)^{\left(\sum_{j=1}^{n} y_j+1\right)}}$$

Although for demonstration purpose we used a flat prior in the above Poisson regression example, a question that arises now is that how the prior for $\tilde{x}$, *in addition to the prior for* θ, should be chosen. Thus, compared to forward problems, which requires prior specification for model parameters θ only, inverse problems have an additional complexity due to the requirement of prior distribution for $\tilde{x}$ in addition to prior requirement for θ. Since θ is the model parameter, the prior for θ, denoted by $\pi(\theta)$, should be chosen independently of $\tilde{x}$. The prior for $\tilde{x}$ may depend upon θ; we denote the joint prior for $(\tilde{x}, \theta)$ by $\pi(\tilde{x}, \theta) = \pi(\theta)\pi(\tilde{x} \mid \theta)$. However, in most cases, it will be convenient to use a prior for $\tilde{x}$, given by $\pi(\tilde{x})$, independent of θ.

In the case of forward problems, possible mis-specification of the prior distribution on model parameters θ is usually investigated by sensitivity analysis. This has been an important element of the philosophies of a number of Bayesians (see, e.g. Berger (1985) and the references therein). Loosely, sensitivity analysis involves trying different reasonable priors and scrutinising the resultant posterior quantities. Re-thinking is necessary if, due to different prior assumptions, the posterior quantities are changed in a way that has practical impact on interpretations or decisions.

The above exposition of inverse problems automatically leads to consider sensitivity analysis for checking possible mis-specification of prior distributions on $\tilde{x}$. For our purposes, it will be assumed that an appropriate prior for θ has been already specified using standard methods for sensitivity analysis using the posteriors of the form $\pi(\theta \mid X, Y) \propto \pi(\theta) \prod_{j=1}^{n} p(y_j \mid x_j, \theta)$ (as an aside, we remark that this posterior has been referred to as 'saturated posterior' by Bhattacharya and Haslett (2007)). For both forward and inverse problems, implementation of sensitivity analysis involves computational challenges. Although the literature contains methodologies to deal with this challenge in the case of forward problems, there exists no literature for corresponding challenges in inverse problems. Hence, in this paper, we will particularly devote ourselves to developing computational techniques for sensitivity analysis for the prior on $\tilde{x}$.

The rest of our paper is structured as follows. In Section 2, we review computational methods available for sensitivity analysis in forward problems. Our proposed method for sensitivity in inverse problems is introduced in Section 3. In Section 4 we demonstrate, with reference to the palaeoclimate model of Haslett *et al.* (2006), that our proposed methodology requires modification when dimensionalities of priors under consideration are different. The modified proposal, which takes into account different dimensionalities of priors, is discussed in Section 5. Armed with the new proposal, in Section 6 we implement sensitivity analysis in the case of the complex palaeoclimate model of Haslett *et al.* (2006), and demonstrate the superiority of our proposal over regular MCMC method. Finally, we conclude in Section 7.

2. COMPUTATIONAL TECHNIQUES FOR SENSITIVITY ANALYSIS IN FORWARD PROBLEMS

For sensitivity analysis in forward problems, where mis-specification of the prior on the model parameters θ needs to be checked, in principle it is possible to compute, via regular MCMC, the posterior $\pi(\theta \mid X, Y)$, for all reasonable priors on θ, denoted by $\pi(\theta)$. Note however, that for many reasonable priors $\pi(\theta)$, this entails recomputing posterior quantities very many times. This could be computationally extremely expensive, particularly for high-dimensional θ. Importance sampling is generally recommended for sensitivity analysis (see, e.g. Athreya *et al.* (1996), Doss (1994)). For a particular prior π_0 (preferably, sufficiently thick-tailed) on the parameter θ, samples $\theta^{(1)}, \ldots, \theta^{(N)}$ are generated from the posterior $\pi_0(\theta \mid X, Y)$, usually by regular MCMC. Then for another prior of interest, π_i ($i = 1, \ldots, M$, say, where M may be large), importance weights of posteriors $\pi_i(\theta \mid X, Y)$ are computed with respect to the 'initial posterior' $\pi_0(\theta \mid X, Y)$. This is given by

$$w_i(\theta) = \pi_i(\theta \mid X, Y)/\pi_0(\theta \mid X, Y) \propto \pi_i(\theta)/\pi_0(\theta) \tag{5}$$

These weights are then used to compute approximations to posterior quantities of $\pi_i(\theta \mid X, Y)$, for example expectation of an appropriate function $h(\theta)$ as

$$\hat{E}_N^{(i)}(h(\theta)) = \frac{1}{N} \sum_{\ell=1}^{N} h(\theta^{(\ell)}) w_i(\theta^{(\ell)}) \tag{6}$$

In fact, if in Equation (6) weights $w_i(\theta)$ are known only up to a constant, rescaling $\hat{E}_N^{(i)}(h(\theta))$ by the average of the weights is sufficient. The quality of the estimate thus obtained depends much on the variability of the importance weights w_i, which, in turn depends upon how close the priors π_i are to the initial prior π_0.

The above technique of sensitivity analysis is appropriate in the case of forward problems. However, such technique is inapplicable in the case of inverse problems, as we demonstrate below.

3. PROPOSED COMPUTATIONAL TECHNIQUES FOR SENSITIVITY ANALYSIS IN INVERSE PROBLEMS

In addition to the set up described in the above Poisson regression problem, suppose that a further set of observations y are available from the model but the corresponding $\tilde{x}$ are unavailable. The interest is to learn about the set of unknown values, $\tilde{x}$; θ is treated as a (possibly multidimensional) nuisance parameter (see, e.g. Berger *et al*. (1999)). Recall that, we assume that an appropriate prior, given by $\pi(\theta)$ is assigned to θ.

Our interest in this case is thus to check sensitivity of the priors on $\tilde{x}$ and not on θ. A convenient way to proceed is to propose priors $\{\pi_0(\cdot \mid \theta),\ \pi_i(\cdot \mid \theta); i = 1, \ldots, M\}$ on $\tilde{x}$, where $\pi_0(\cdot \mid \theta)$ is the prior of main interest and $\{\pi_i(\cdot \mid \theta);\ i = 1, \ldots, M\}$ are considered variations of the former. The interest is then to check sensitivity with respect to the posterior $\pi_i(\tilde{x} \mid X, Y, y) = \int \pi_i(\tilde{x}, \theta \mid X, Y, y)\mathrm{d}\theta$. However, since the functional form of the posteriors involve integrating out the nuisance parameter θ, this may not be available, even up to a constant. Hence importance weights, given by

$$\begin{aligned} w_i(\tilde{x}) &= \pi_i(\tilde{x} \mid X, Y, y)/\pi_0(\tilde{x} \mid X, Y, y) \\ &= \frac{\int \pi_i(\tilde{x}, \theta \mid X, Y, y)\mathrm{d}\theta}{\int \pi_0(\tilde{x}, \theta \mid X, Y, y)\mathrm{d}\theta} \end{aligned}$$

will not be available here, unlike in the forward case, which was given by Equation (5).

In principle, it is possible to apply IRMCMC of Bhattacharya and Haslett (2007) to such inverse problems. Only appropriate modification of their algorithm is necessary, which we discuss below. Put simply, one can realise from $\pi_0(\tilde{x}, \theta \mid X, Y, y)$, typically by regular MCMC, a sample of $(\tilde{x}, \theta)$. Then using importance weights of the form given by

$$\begin{aligned} w_{0,i}(\tilde{x}, \theta) &= \frac{\pi_i(\tilde{x}, \theta \mid X, Y, y)}{\pi_0(\tilde{x}, \theta \mid X, Y, y)} \\ &\propto \frac{\pi_i(\tilde{x} \mid \theta)}{\pi_0(\tilde{x} \mid \theta)} \end{aligned} \tag{7}$$

one can realise a subsample of θ realisations corresponding to the posterior $\pi_i(\theta \mid X, Y, y)$. Given each realised θ, realisations of $\tilde{x}$ can be obtained from $\pi_i(\tilde{x} \mid y, \theta)$, typically by regular MCMC. The realised $\tilde{x}$ can then be said to be samples from the target posterior $\pi_i(\tilde{x} \mid X, Y, y)$. This method can be repeated for each $i = 1, \ldots, M$. Note that the essence of this proposal is exactly the IRMCMC proposal as described in Bhattacharya and Haslett (2007) in the context of cross-validation. In the context of sensitivity analysis the proposal of Bhattacharya and Haslett (2007) can be modified as follows.

1. Choose an initial prior $\pi_0(\tilde{x} \mid \theta)$. Use the initial posterior $\pi_0(\tilde{x}, \theta \mid X, Y, y) \propto \pi(\theta)\pi(\tilde{x} \mid \theta)\pi(y \mid \tilde{x}, \theta) \prod_{j=1}^{n} p(y_j \mid x_j, \theta)$ as the importance sampling density.
2. From this density, sample values $(\tilde{x}^{(\ell)}, \theta^{(\ell)})$; $\ell = 1, \ldots, N$, for large N. Typically, regular MCMC will be used for sampling.
3. For $i \in \{1, \ldots, M\}$ do
   a. For each sample value $(\tilde{x}^{(\ell)}, \theta^{(\ell)})$, compute importance weights $w_{0,i}^{(\ell)} = w_{0,i}(\tilde{x}^{(\ell)}, \theta^{(\ell)})$, where the importance weight function is given by

$$w_{0,i}(\tilde{x}, \theta) \propto \frac{\pi_i(\tilde{x} \mid \theta)}{\pi_0(\tilde{x} \mid \theta)} \tag{8}$$

   b. For $k \in \{1, \ldots, K\}$
      (i) Sample $\tilde{\theta}^{(k)}$ from $\theta^{(1)}, \ldots, \theta^{(N)}$ where the probability of sampling $\theta^{(\ell)}$ is proportional to $w_{0,i}^{(\ell)}$.
      (ii) For *fixed* $\theta = \tilde{\theta}^{(k)}$, draw T times from $\pi_i(\tilde{x} \mid y, \tilde{\theta}^{(k)})$. Thus, for the Poisson regression case, with prior π_i on $\tilde{x}$ (and assuming that the prior on $\tilde{x}$ is independent of θ),

$$\pi_i(\tilde{x} \mid y, \theta) \propto \pi_i(\tilde{x}) \exp(-\theta\tilde{x})\tilde{x}^y \tag{9}$$

      If π_i is a flat prior, then the above distribution (in Eq. 9) is the Gamma distribution. Note that in general it is not easy to sample from $\pi_i(\tilde{x} \mid y_i, \theta)$, even when $\tilde{x}$ is univariate, and we recommend MCMC for generality. For example for the Poisson regression case, if the prior $\pi_i(\tilde{x})$ is given by a Cauchy distribution, truncated on $(0, \infty)$, then

$$\pi_i(\tilde{x} \mid y, \theta) \propto \frac{1}{1 + \tilde{x}^2} \exp(-\theta\tilde{x})\tilde{x}^y \tag{10}$$

      To generate samples, from Equation (10), MCMC seems to be the simplest methodology. Here we draw attention of the reader to an interesting parallel implementation issue. Note that, for each different k, a separate MCMC, with separate starting value, is required to draw samples from $\pi_i(\tilde{x} \mid y, \tilde{\theta}^{(k)})$. Hence, MCMC runs needed to explore $\pi_i(\tilde{x} \mid y, \tilde{\theta}^{(k)})$ for each k are independent of each other. Because of this independence, the MCMC computations can be done in separate parallel processors. This is not only computationally efficient, but the separate starting values of the MCMC algorithms for each k enable exploration of the posterior $\pi_i(\tilde{x} \mid X, Y, y)$ much more reliably; in particular, this independence makes the algorithm very suitable for adequate exploration of multimodal posteriors. For details, see Bhattacharya and Haslett (2007).
   c. Store as $\{\tilde{x}_i^{(1)}, \ldots, \tilde{x}_i^{(KT)}\}$ the $K \times T$ draws of $\tilde{x}$ corresponding to prior π_i as representative of the posterior $\pi_i(\tilde{x} \mid X, Y, y)$.

An important technical question is whether IR of $\{\theta^{(\ell)}; \ell = 1, \ldots, N\}$ should be used with or without replacement. Indeed, most of the references to IR in the literature use sampling with replacement. See, for example Gelfand *et al.* (1992), Newton and Raftery (1994), O'Hagan and Forster (2004). However, Gelman *et al*. (1995), Stern and Cressie (2000) recommend IR without replacement. It has been argued that sampling without replacement can provide protection against highly variable importance weights. Recently, Skare *et al*. (2003) formally prove a theorem that IR without replacement is better than IR with replacement, with respect to the total variation norm. Thus, in our proposal we recommend the former. Bhattacharya (2004) provides further details in this context including a comparison of IR with/without replacement.

Thus IRMCMC may be applied in a straightforward manner to carry out sensitivity analysis in inverse problems. There is no issue now as to the choice of importance sampling distribution: it is the posterior $\pi_0(\tilde{x}, \theta \mid X, Y, y)$ computed with reference to the prior $\pi_0(\cdot \mid \theta)$. Nothing more needs to be said.

There is however a variation, which deserves careful attention. For it is not necessarily the case that the priors $\pi_0(\tilde{x} \mid \theta)$ and $\pi_i(\tilde{x} \mid \theta)$ are of the same dimension, and hence further modification of the above algorithm is necessary. We demonstrate this next with the palaeoclimate model of Haslett *et al*. (2006).

4. A PALAEOCLIMATE MODEL

Haslett *et al*. (2006) use 'modern' observed training data on pollen and two climate variables, GDD5 (growing degree days above 5°C and MTCO (mean temperature of the coldest month) and attempt to reconstruct, using 'fossil' pollen data, as well as the observed modern training data, the two unobserved climates that prevailed over Glendalough in Ireland about 15,000 calendar years ago. Thus, the inverse nature of the problem is already apparent.

The compositional pollen data arise from counts. Each vector of proportions represents each of a set of distinguishable taxa in a sample extracted from the sediment. For the purposes of their study, Haslett *et al*. (2006) use $m = 14$ different taxa; this yields a vector y_i of counts, with elements y_{ik} for $k = 1, \ldots, m$. For this problem, 7815 such vectors of counts are available.

In Haslett *et al*. (2006), for each sample, modern or fossil, $y_i \mid \mathbf{p}_i, n_i \sim Multinomial(n_i, \mathbf{p}_i)$ independently; here $\mathbf{p}_i$ denotes the underlying composition of the pollen assemblage in the sediment sample and n_i the total count. The elements p_{ik} refer to the kth taxon at the ith site.

The Dirichlet mixture of multinomials provides a natural and convenient model for such variation, being conjugate to the multinomial. After integrating out $\mathbf{p}_i$, the marginal distribution of y_i is obtained as the compound multinomial distribution (see Dey and Maiti (2002)), given by

$$\pi(y_i \mid \gamma_i, \delta, n_i) = \frac{n_i! \Gamma(\delta)}{\Gamma(n_i + \delta)} \prod_{k=1}^{m} \left(\frac{\Gamma(y_{ik} + \delta\gamma_{ik})}{\Gamma(\delta\gamma_{ik}) y_{ik}!} \right) \tag{11}$$

Under this parameterisation, $E(p_{ik} \mid \gamma_{ik}, \delta, n_i) = \gamma_{ik}$. The δ parameter has a simple interpretation as controlling 'extra-multinomial' dispersion. Note that in this formulation $\sum_k \gamma_{ik} = 1$. Haslett *et al*. (2006) model the γ_{ik} as functions in a real two-dimensional climate space as $\gamma(x_i, \phi_k)$ to relate the climate x_i to the propensity of the pollen from the kth taxon to occur in lake sediments. The two dimensional climate space used in this case is shown in Figure 1. The climate space consists of 778 points, each associated with a 13-dimensional random variable. Thus, there are $778 \times 13 = 10\,114$ parameters associated with the climate space. We denote by θ the entire set of parameters γ_{ik} and δ. Note that γ_{ik} depend directly upon observed training data and hence should be interpreted as latent variables, rather than parameters. The model for γ_{ik} does not depend upon any unknown parameters.

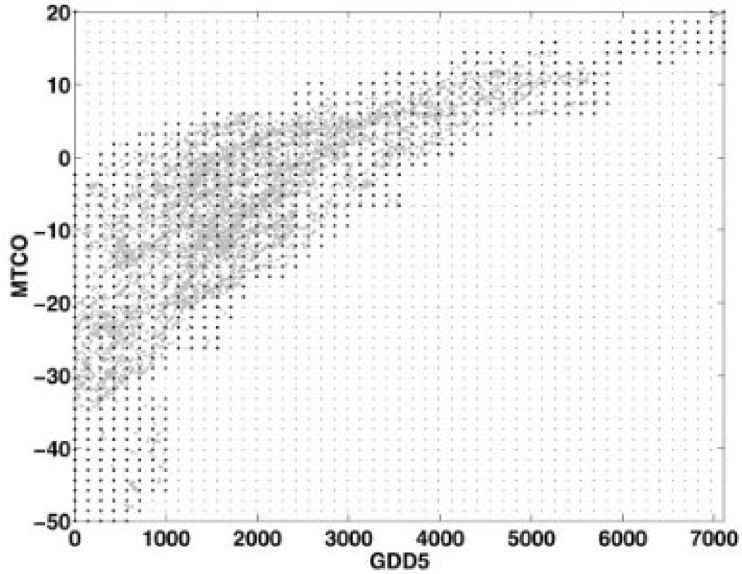

Figure 1. Support lattice for the response surfaces. The bold black points stand for the lattice points that form the support of the response surfaces; the light black points denote the regular lattice, and the grey points denote the training data set

Hence there is no issue regarding prior distribution of γ_{ik}. However, δ must be regarded as an unknown parameter, and in the absence of any information regarding δ, a non-informative prior has been assumed for this parameter. So, there is no requirement for conducting sensitivity analysis of the prior on δ either. For explicit details on the modelling of response surfaces see Haslett *et al.* (2006).

4.1. Temporal smoothness

In Haslett *et al.* (2006) the modern pollen samples Y and the modern climate values X are known but the prehistoric climates, denoted here by $\tilde{x}$, are unknown, although the corresponding fossil pollen samples, which we denote by y, are known. Observe that, each x_j of X, as well as each component of $\tilde{x}$, is bivariate, representing two climate variables, GDD5 and MTCO. The bivariate components of $\tilde{x}$ may be more readily denoted as $\tilde{x}(t_j)$, where the t_j are in radiocarbon years before present (RCYBP). In this palaeoclimate model, $j = 1$ denotes the deepest (oldest) sample and $j = 150$ the most recent.

Climate change exhibits some degree of smoothness in time. Loosely speaking, climate changes can be characterised as 'small,' mostly, but occasionally very large. Haslett *et al.* (2006) model this smoothness stochastically by specifying an appropriate family of priors. Light can be shed upon this by an examination of the ice core data shown in Figure 2. This has been obtained from the stable oxygen isotope $\delta^{18}O$ data from the ice of the GISP2 core drilled near the summit of the Greenland ice sheet (see, e.g. Stuiver *et al.* (1995)). This provides a basis for prior information regarding the temporal properties of the climate system of Glendalough in Ireland although the data pertains to temperature over Greenland, which is some distance away from Ireland. The variation in Figure 2 can be adequately modelled by a random walk. But the normal scores plot of the increments strongly suggests that the variation is much longer tailed than the normal distribution. In fact these increments are quite well described by the t-distribution with 8 degrees of freedom. Based on exploratory analysis of the Greenland ice core data, Haslett *et al.*

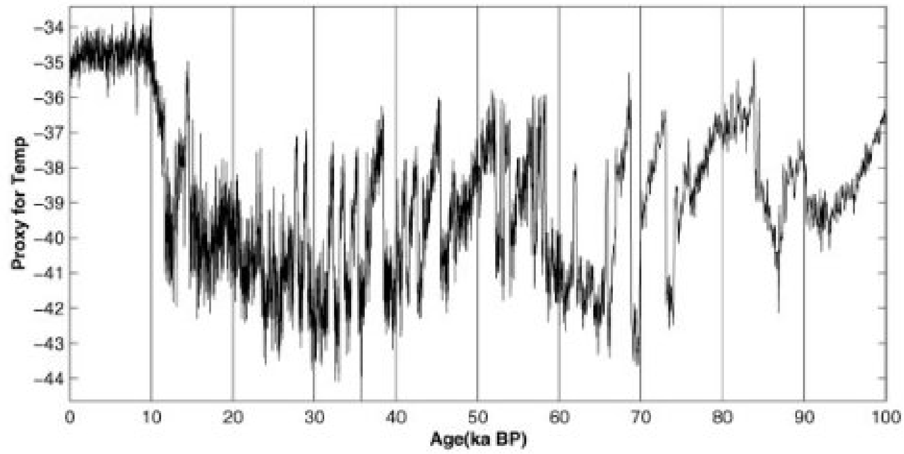

Figure 2. Greenland ice core data. Note that the ages are given in 1000 calendar years before present (ka BP)

(2006) consider several different priors for the dependence between the unknown $\tilde{x}(t_j)$. In particular, they assume that each is of the form $\tilde{x}(t_j) = \tilde{x}(t_{j-1}) + A(t_j)\varepsilon(t_j)$, where $A_j = A(t_j)$ is a matrix; recall that climate has two dimensions in this case. The distributions of the innovations $(\varepsilon_{j1}, \varepsilon_{j2})$ are what distinguish different priors. They compute the posterior distributions of reconstructed climates corresponding to several choices of priors on the innovations - Student's t distribution with 8 degrees of freedom, the Cauchy distribution and the Gaussian distribution. They also consider a prior with complete temporal independence, that is the prior here is simply uniform on the climate space with climates at different times being independent of each other. In all cases, GDD5 and MTCO are assumed to be distributed independently of each other. It is further assumed that $A_j = |t_j - t_{j-1}|\Delta$ where Δ is diagonal with $(\Delta_{\mathrm{GDD5}}, \Delta_{\mathrm{MTCO}})$ being the diagonal elements. Thus we may look upon Δ as a two-dimensional random variable.

Formally, for priors with temporal smoothness,

$$\pi(\tilde{x} \mid \Delta) = \pi_{\mathrm{GDD5}}(\tilde{x} \mid \Delta)\pi_{\mathrm{MTCO}}(x \mid \Delta) \tag{12}$$

Specifically, using the random walk model it can be written as

$$\pi_{\mathrm{GDD5}}(\tilde{x}|\Delta) = \prod_{j=2,150} g_d(\tilde{x}_{\mathrm{GDD5}}(t_j) \mid \mu_{\mathrm{GDD5}}(t_j, t_{j-1}), \sigma_{\mathrm{GDD5}}(t_j, t_{j-1}))$$

and similarly for MTCO. By $g_d(u \mid \mu, \sigma)$ we denote the value of the pdf of a t distribution with d degrees of freedom, evaluated at $\sigma^{-1}(u - \mu)$; we set $\mu = x_{\mathrm{GDD5}}(t_{j-1})$ and $\sigma^2 = \Delta^2_{\mathrm{GDD5}}(t_j - t_{j-1})^2$. Note that the Cauchy is a t distribution with one degree of freedom and Gaussian is a t distribution with infinite degrees of freedom.

Thus, Haslett *et al*. (2006) consider a prior with complete temporal independence, which we denote by π_0, in addition to the following three random walk priors:

(1) Gaussian random walk prior, denoted by π_1.
(2) Student's t random walk prior with $d = 8$ degrees of freedom, denoted by π_2.
(3) Cauchy random walk prior, denoted by π_3.

For sensitivity analysis in this inverse problem, it is required to compute, with respect to the priors $\{\pi_i; i = 0, 1, 2, 3\}$, the posteriors $\{\pi_i(\tilde{x} \mid X, Y, y); i = 0, 1, 2, 3\}$.

4.2. *Technical difficulty associated with implementation of sensitivity analysis*

In Section 3, we have demonstrated how a modification of IRMCMC of Bhattacharya and Haslett (2007) can be used as an efficient method for conducting sensitivity analysis in inverse problems. In this particular problem, the proposal entails first simulation of $(\tilde{x}^{(\ell)}, \Delta^{(\ell)}, \theta^{(\ell)}); \ell = 1, \ldots, N$, for large N using regular MCMC, from the posterior $\pi_0(\tilde{x}, \Delta, \theta \mid X, Y, y)$. The latter corresponds to a specific temporal smoothness prior, denoted by π_0. It is then necessary to reweigh towards posteriors corresponding to the remaining temporal smoothness priors π_i, using importance weights proportional to $\pi_i(\tilde{x})/\pi_0(\tilde{x})$ to obtain $\{\tilde{\theta}^{(k)}; k = 1, \ldots, K\}$. Regular MCMC may then be used for the simulation of climate variables and Δ from $\pi_i(\tilde{x}, \Delta \mid y, \tilde{\theta}^{(k)})$, and store the subsequent samples of $\tilde{x}$ corresponding to the ith prior π_i as $\{\tilde{x}_i^{(1)}, \ldots, \tilde{x}_i^{(KT)}\}$.

However, the idea of reweighing the samples of θ using appropriate importance weights is rendered invalid when, among the possible priors π_i on climate, the prior π_0, with no temporal smoothness, is also considered. Observe that π_0 does not consist of the two-dimensional random variable Δ and so has dimension two less than the other priors π_1, π_2, π_3, which involve temporal dependence. In other words, the joint posterior of the parameters, with respect to π_0 is $\pi_0(\tilde{x}, \theta \mid X, Y, y)$; but with respect to the priors $\{\pi_i; i = 1, 2, 3\}$, the corresponding posteriors are $\pi_i(\tilde{x}, \theta, \Delta \mid X, Y, y); i = 1, 2, 3$. Thus, the dimensionality of the posterior $\pi_0(\tilde{x}, \theta \mid X, Y, y)$ is two less than those of $\pi_i(\tilde{x}, \theta, \Delta \mid X, Y, y); i = 1, 2, 3$. Since the dimensions are different, it is not possible to compute importance weights, which, depending upon the choice of importance sampling density, must be of the form $w_{0,i} \propto \pi_i(\tilde{x}, \theta, \Delta \mid X, Y, y)/\pi_0(\tilde{x}, \theta \mid X, Y, y)$ or $1/w_{0,i}$, for $i = 1, 2, 3$.

5. PROPOSAL FOR IMPLEMENTATION OF SENSITIVITY ANALYSIS WHEN PRIORS HAVE DIFFERENT DIMENSIONALITIES

Our proposal to handle sensitivity analysis in the case of priors with different dimensionalities is to generate a sample of θ realisations from some distribution of θ that adequately approximates all the target posteriors of θ given by $\pi_i(\theta \mid X, Y, y) = \int \pi_i(\tilde{x}, \theta \mid X, Y, y)\mathrm{d}\tilde{x}$. We argue below that the saturated posterior $\pi(\theta \mid X, Y)$ is a good candidate for approximating $\pi_i(\theta \mid X, Y, y)$. We can then resample from the available sample of θ-realisations simply by random sampling, *without computing any importance weights*. We thus consider a special case of IRMCMC described in Section 3 where the weights are equal. This solves the dimensionality problem since we no longer need to compute importance weights. Here are the details.

The posterior of the unknown fossil climates $\tilde{x}$ with temporal smoothness priors $\pi_i; i = 1, 2, 3$ can be written as

$$\begin{aligned}
\pi_i(\tilde{x} \mid X, Y, y) &= \int \pi_i(\tilde{x}, \Delta, \theta \mid X, Y, y) d\Delta \mathrm{d}\theta \\
&= \int \pi_i(\tilde{x}, \Delta \mid \theta, y)\pi_i(\theta \mid X, Y, y) d\Delta \mathrm{d}\theta \\
&\approx \int \pi_i(\tilde{x}, \Delta \mid \theta, y)\pi(\theta \mid X, Y) d\Delta \mathrm{d}\theta \qquad (13)
\end{aligned}$$

In the case where no temporal smoothness is assumed (that is, with prior π_0), the posterior of $\tilde{x}$ is given by

$$\begin{aligned}\pi_0(\tilde{x} \mid X, Y, y) &= \int \pi_0(\tilde{x}, \theta \mid X, Y, y)\mathrm{d}\theta \\ &= \int \pi_0(\tilde{x} \mid \theta, y)\pi_0(\theta \mid X, Y, y)\mathrm{d}\theta \\ &\approx \int \pi_0(\tilde{x} \mid \theta, y)\pi(\theta \mid X, Y)\mathrm{d}\theta \qquad (14)\end{aligned}$$

In both Equations (13) and (14), we assumed that the approximation $\pi_i(\theta \mid X, Y, y) \approx \pi(\theta \mid X, Y)$ holds. For large data size, as here (recall 7815 vectors of pollen counts are available), this is indeed a valid assumption, as shown by the theoretical exposition in Chapter 5 of Bhattacharya (2004). We first generate (typically by regular MCMC) and store a sufficiently large sample of θ from the saturated posterior $\pi(\theta \mid X, Y)$. Then, for fixed θ, we simulate $(\tilde{x}, \Delta)$ from $\pi_i(\tilde{x}, \Delta \mid \theta, y)$ in the cases of priors with temporal dependence prior π_i, and store only realisations corresponding to $\tilde{x}$. In the case of the independence prior π_0, we generate (and store) $\tilde{x}$ from $\pi_0(\tilde{x}, \Delta \mid \theta, y)$. This is certainly the basic idea of IRMCMC as described in Section 3. The only difference is that, since importance weights are unavailable in this case, we rely on asymptotics instead of exact computation using IR, assuming implicitly that the importance weights are all equal. In other words, simple random sampling of the realisations of θ obtained from the saturated posterior may serve as an easily available alternative to the more precise IR, which is unavailable.

In the next section, we compare results of sensitivity analysis obtained by IRMCMC and regular MCMC in the case of the palaeoclimate problem of Haslett *et al.* (2006).

6. RESULTS OF SENSITIVITY ANALYSIS IN THE CASE OF THE COMPLEX PALAEOCLIMATE PROBLEM

We implement sensitivity analysis in this inverse palaeoclimate problem by two methods: regular MCMC and our proposed IRMCMC method. With regular MCMC, after assessing the approximate convergence time as 1000 iterations, we stored the next 10 000 realisations of climate variables for each of the four posteriors corresponding to the four priors $\pi_0, \pi_1, \pi_2, \pi_3$.

For IRMCMC implementation, we first obtained 10 000 realisations by regular MCMC from the saturated posterior $\pi(\theta \mid X, Y)$, after discarding the first 1000 realisations as burn-in. Then, from the sample of 10 000 realisations of θ we randomly selected 1000 values. In other words, we took N=10 000 and $K = 1000$. For each of the randomly selected 1000 realisations of θ, we obtained 10 MCMC realisations of climate from $\pi_i(\tilde{x} \mid y, \theta)$, where MCMC for a new θ is started afresh with a new initial value. So, for each of the four posteriors we obtained by IRMCMC KT=10 000 realisations.

The real advantage of using IRMCMC is the computational speed. The total time taken to compute the posteriors corresponding to all four priors using four regular MCMC runs took about 22 h in a parallel computing environment using 8 processors. But with our proposed IRMCMC-based method, a single regular MCMC run for the saturated posterior $\pi(\theta \mid X, Y)$ was needed. Random sampling from the samples of θ and simulating large samples from posteriors given fixed θ were done at almost no computational cost. In our case, computation of the saturated posterior took about 5 h 28 min and then

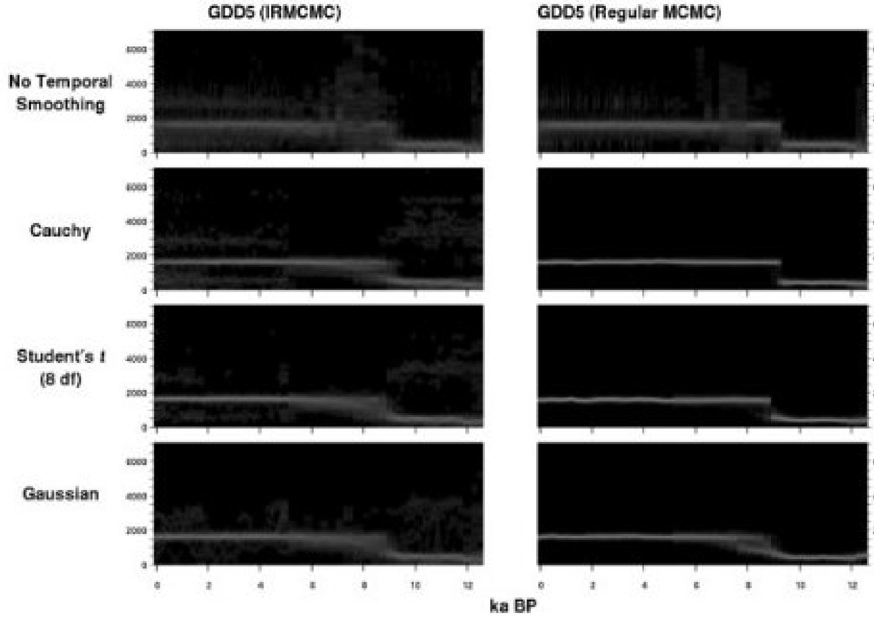

Figure 3. Reconstructions of GDD5 using IRMCMC and regular MCMC

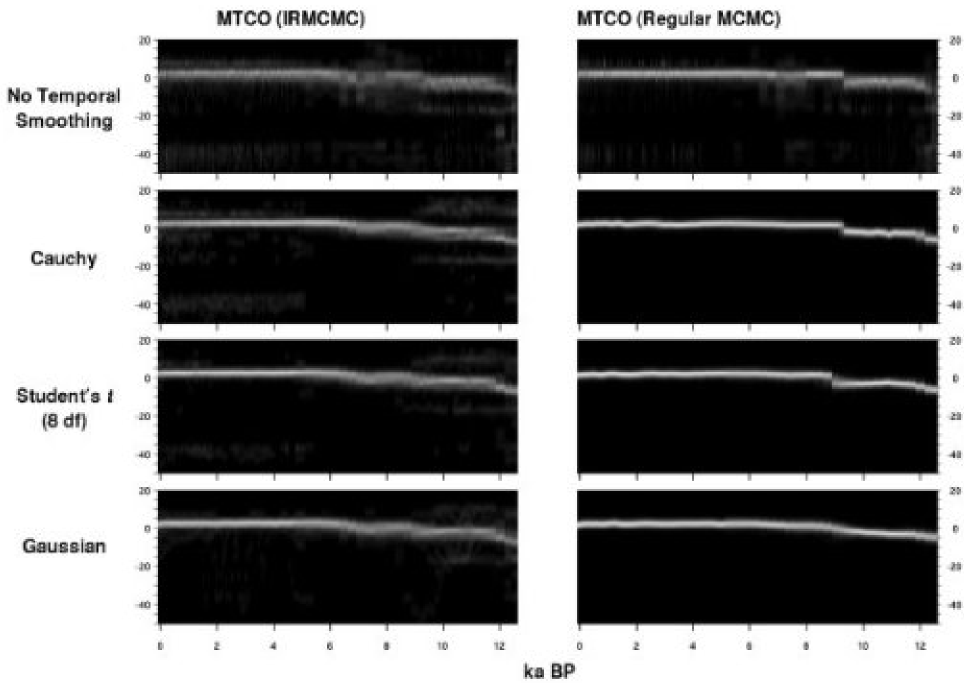

Figure 4. Reconstructions of MTCO using IRMCMC and regular MCMC

using the realised values of θ the remaining computation took just about 12 min. All computations have been carried out in a parallel computing environment using 8 processors.

The results of sensitivity analysis using IRMCMC and regular MCMC are shown in Figures 3 and 4 respectively. The posteriors under the four different priors considered are broadly in agreement with each other; however, there seems to be an issue regarding multimodalities of the posteriors. It is clear that the posteriors under all four priors exhibit multimodalities, but those corresponding to the temporal independence prior π_0 exhibits most variability, particularly, in terms of the number of minor modes. Actually, the multimodalities are to be expected in pollen-based palaeoclimate reconstruction problems. Indeed, Haslett *et al*. (2006) remark that a source of the technical difficulty they encountered, is that of the multimodal nature of the posterior. This multimodality arises naturally in palaeoclimate studies since the pollen species have different (and/or more than one) climate preferences; this sends conflicting signals to the posterior forcing it to be multimodal. For detailed discussion on multimodality in palaleoclimate studies, see Bhattacharya and Haslett (2007), Bhattacharya and Haslett (2004), Haslett *et al*. (2006).

The figures further show that posteriors obtained by IRMCMC exhibit more multimodalities than those obtained by regular MCMC. As already argued in Section 3, due to independence of realisations of θ, IRMCMC explores the solution space more reliably than regular MCMC and hence results obtained by IRMCMC are more reliable. Similar conclusion has been drawn by Bhattacharya and Haslett (2007), who demonstrate in detail, in the case of a palaeoclimate problem, that IRMCMC is much better suited than regular MCMC for exploring multimodal distributions.

7. CONCLUSION

In this paper, we have discussed that sensitivity analysis in inverse problems requires more careful attention compared to sensitivity analysis in forward problems. We have presented a very fast computational methodology which is based on IRMCMC of Bhattacharya and Haslett (2007), and have demonstrated, with application to the complex palaeoclimate model of Haslett *et al*. (2006), that it can deal very effectively with sensitivity analysis in inverse problems. Furthermore, we have demonstrated the superiority of our proposed method over regular MCMC methods. In particular, we have demonstrated that, compared to regular MCMC methods our proposed method is highly suitable for exploration of multimodal posteriors, and is also many times more faster.

However, in this paper, we did not concern ourselves with philosophical aspects of sensitivity analysis in inverse problems. This may be an interesting topic for future work.

ACKNOWLEDGEMENTS

The work was done when the author was pursuing his PhD at Trinity College Dublin and was supported by Enterprise Ireland 14 grant SC/2001/171 for palaeoclimate reconstruction. The author is grateful to Professor J. Haslett for advice on presentation of the paper. The figures are obtained using Matlab codes, which are modified versions of basic initial codes written by Matt Whiley. The data sets used in this paper can be obtained from http://www.rss.org.uk/preprints.

REFERENCES

Athreya KB, Doss H, Sethuraman J. 1996. On the convergence of the Markov chain simulation method. *The Annals of Statistics* **24**(1): 69–100.

Berger JO. 1985. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag: New York.

Berger JO, Liseo B, Wolpert RL. 1999. Integrated likelihood methods for eliminating nuisance parameters (with discussion). *Statistical Science* **14**(1): 1–28.

Bhattacharya S. 2004. *Importance Resampling MCMC: A Methodology for Cross-Validation in Inverse Problems and Its Applications in Model Assessment*. Doctoral thesis, Department of Statistics, Trinity College Dublin. Available at http://www.tcd.ie/Statistics/JHpersonal/thesis.pdf

Bhattacharya S. 2006. A Bayesian semiparametric model for organism based environmental reconstruction. *Environmetrics* **17**(7): 763–776.

Bhattacharya S, Haslett J. 2004. Fast Cross Validation of a Palaeoclimate Model using IRMCMC. Technical report, Trinity College, Dublin, Ireland. Available at http://www.tcd.ie/Statistics/JHpersonal/research.htm

Bhattacharya S, Haslett J. 2007. Importance re-sampling MCMC for cross-validation in inverse problems. *Bayesian Analysis*. To appear.

Cornford D, Csató L, Evans DJ, Opper M. 2004. Bayesian analysis of the scatterometer wind retrieval inverse problem: some new approaches (with discussion). *Journal of the Royal Statistical Society Series B* **66**(3): 609–626.

Dey D, Maiti T. 2002. Dirichlet multinomial distribution. In *Encyclopedia of Environmetrics*. El-Shaarawi A and Piegorsch W (eds). Fisher: Stuttgart. Wiley; 522–523.

Doss H. 1994. Discussion: Markov chains for exploring posterior distributions. *Annals of Statistics* **22**(4): 1728–1734.

Gelfand AE, Dey DK, Chang H. 1992. Model determination using predictive distributions with implementation via sampling-based methods (with discussion). In *Bayesian Statistics*, vol. 4 Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds). Oxford University Press: Oxford; 147–167.

Gelman A, Carlin JB, Stern HS, Rubin DR. 2004. *Bayesian Data Analysis*, (2nd edn). Chapman and Hall: Boca Raton.

Haario H, Laine M, Lehtinen M, Saksman E, Tamminen J. 2004. Markov chain Monte Carlo methods for high dimensional inversion in remote sensing (with discussion). *Journal of the Royal Statistical Society Series B* **66**(3): 591–607.

Haslett J, Whiley M, Bhattacharya S, Salter-Townshend M, Wilson SP, Allen JRM, Huntley B, Mitchell FJG. 2006. Bayesian palaeoclimate reconstruction (with discussion). *Journal of the Royal Statistical Society Series A* **169**(3): 395–438.

Newton MA, Raftery AE. 1994. Approximate Bayesian inference with the weighted likelihood bootstrap (with discussion). *Journal of the Royal Statistical Society Series B* **56**(1): 3–48.

O'Hagan A, Forster J. 2004. *Kendall's Advanced Theory of Statistics, Bayesian Inference* (Vol. 2B). Arnold: London.

Rubin D. 1988. Using the SIR algorithm to simulate posterior distributions. In *Bayesian Statistics*, vol. 3 Bernardo JM, DeGroot MH, Lindley DV, Smith AFM (eds). Oxford: New York; 395–402.

Skare Ø, Bølviken E, Holden L. 2003. Improved sampling-importance resampling and reduced bias importance sampling. *Scandivanian Journal of Statistics* **30**: 719–737.

Stern HS, Cressie N. 2000. Posterior predictive model checks for disease mapping models. *Statistics in Medicine* **19**: 2377–2397.

Stuiver M, Grootes PM, Braziunas TF. 1995. The GISP2 δ^{18}O climate record of the past 16,500 years and the role of the sun, ocean, and volcanoes. *Quaternary Research* **44**: 341–354.

Tierney L. 1994. Markov Chains for exploring posterior distributions (with discussion). *Annals of Statistics* **22**(4): 1701–1702.

Vasko K, Toivonen HT, Korhola A. 2000. A Bayesian multinomial Gaussian response model for organism based environmental reconstruction. *Journal of Paleolimnology* **24**: 243–250.