On the Relationship Between Markov Chain Monte Carlo Methods for Model Uncertainty

Simon J. GODSILL

This article considers Markov chain computational methods for incorporating uncertainty about the *dimension* of a parameter when performing inference within a Bayesian setting. A general class of methods is proposed for performing such computations, based upon a product space representation of the problem which is similar to that of Carlin and Chib. It is shown that all of the existing algorithms for incorporation of model uncertainty into Markov chain Monte Carlo (MCMC) can be derived as special cases of this general class of methods. In particular, we show that the popular reversible jump method is obtained when a special form of Metropolis–Hastings (M–H) algorithm is applied to the product space. Furthermore, the Gibbs sampling method and the variable selection method are shown to derive straightforwardly from the general framework. We believe that these new relationships between methods, which were until now seen as diverse procedures, are an important aid to the understanding of MCMC model selection procedures and may assist in the future development of improved procedures. Our discussion also sheds some light upon the important issues of "pseudo-prior" selection in the case of the Carlin and Chib sampler and choice of proposal distribution in the case of reversible jump. Finally, we propose efficient reversible jump proposal schemes that take advantage of any analytic structure that may be present in the model. These proposal schemes are compared with a standard reversible jump scheme for the problem of model order uncertainty in autoregressive time series, demonstrating the improvements which can be achieved through careful choice of proposals.

Key Words: Bayes; Jump diffusion; Model selection; Reversible jump; Variable selection.

1. INTRODUCTION

1.1 BAYESIAN MODEL UNCERTAINTY

Within a Bayesian setting model uncertainty can be handled in a parametric fashion through the use of posterior model probabilities. Suppose there exist M candidate models, one of which is assumed to be a perfect statistical description of an observed data vector y.

Simon Godsill is a University Lecturer in Information Engineering, Signal Processing Group, Engineering Department, University of Cambridge, CB2 1PZ, (E-mail: sjg@eng.cam.ac.uk).

Journal of Computational and Graphical Statistics, Volume 10, Number 2, Pages 1–19

Associated with each model is a likelihood $p(y|\theta_k, k)$ that depends upon an (unknown) set of parameters θ_k, where $k \in \{1, \ldots, M\}$ denotes the kth model in the list of candidates. In general θ_k may be multivariate and may have different dimensionality and support Θ_k in different models. A prior distribution $p(\theta_k|k)$ is assigned to each parameter vector and a prior distribution $p(k)$ to the model number, reflecting prior knowledge about the probabilities of individual models. The posterior model probability for model k is then obtained as

$$p(k|y) = \frac{p(y|k)p(k)}{p(y)} = \frac{\int_{\Theta_k} p(y|\theta_k, k)p(\theta_k|k)d\theta_k \, p(k)}{p(y)}. \tag{1.1}$$

The term $p(y|k)$ is sometimes referred to as the *marginal likelihood* for model k. We assume throughout that the parameter priors $p(\theta_k|k)$ are proper. In some cases the goal of the statistical analysis may simply be to summarize the relative posterior probabilities of the individual models or to estimate a single "best" model through the use of some suitable risk function. In many applied scenarios, however, model uncertainty can be incorporated into tasks such as forecasting, interpolation, smoothing, or signal extraction (West and Harrison 1997) through use of "model mixing," in which model-dependent inferences are combined together by weighting according to their posterior probabilities (Hoeting, Raftery, and Madigan 1996; Raftery, Madigan, and Hoeting 1997).

1.2 MCMC METHODS FOR MODEL UNCERTAINTY

Calculation of posterior model probabilities is rarely achievable in closed form for realistic models. Approximation methods may be used, and there is a large array of tools available (see, e.g., Raftery 1996 for a good review). Another effective means of achieving this is through a Monte Carlo sampling scheme. For distributions of parameters with fixed dimensionality a suitable scheme for drawing a dependent sequence of samples from the joint posterior is Markov chain Monte Carlo (MCMC). MCMC methods (Metropolis et al. 1953; Geman and Geman 1984; Gelfand and Smith 1990; Hastings 1970) have become well established over recent years as a powerful computational tool for analysis of complex statistical problems. Until relatively recently, however, these methods were applied in statistics only to problems with fixed dimensionality.

A direct MCMC approach to the solution of the variable dimension problem is to estimate posterior model probabilities from independent MCMC chains running in each model. This is the approach of, for example, Chib (1995) and Chib and Greenberg (1998). An appealing alternative is to perform MCMC simulation over both model parameters and model number: if one can draw random samples (θ_k^i, k^i) from the joint posterior distribution $p(\theta_k, k|y)$, then Monte Carlo estimates can readily be made for any required posterior quantities. It is then hoped that models with insignificant probability are visited only rarely, while the majority of the computational effort is expended in exploration of models with high probability. This article considers only computational schemes of this latter variety, since we believe that these offer greater potential in the solution of the complex modeling requirements of many realistic applied problems which have high-dimensional parameter spaces and many competing models. However, we note that in cases where there is no obvious relationship—such as a nested structure between the parameters of competing

models—the direct MCMC approaches are likely to be the method of choice at the current time.

Currently, the most flexible and popular MCMC model sampling scheme is the reversible jump sampler (Green 1995). In this scheme, which is covered more fully in Section 2.3, a modified version of the Metropolis–Hastings method is developed based on the detailed balance condition for the distribution $p(\theta_k, k|y)$. Reversible jump can be viewed as a generalization of the earlier jump diffusion methods (Grenander and Miller 1991, 1994; Phillips and Smith 1994).

An alternative scheme, that of Carlin and Chib (1995), develops a product space for all possible model parameters and a model indexing variable. The space has constant dimensionality and hence a Gibbs sampler may be applied directly without concern for the varying dimension aspects of the model uncertainty problem. The technique has an appealing simplicity but is not easily applicable to problems with more than a handful of competing models, owing to the necessity of choosing and generating from a large number of so-called "pseudo-priors." Other more specialized model space samplers include the stochastic search variable selection (SSVS) methods (George and McCulloch 1993, 1996; Geweke 1996; Kuo and Mallick 1998) and the MCMC model combination (MC^3) methods of Madigan and York (1995), developed in the special context of decomposable graphical models.

The field of MCMC model uncertainty is rapidly growing in application. A few of the many applications of reversible jump methods can be found in Richardson and Green (1997); Denison, Mallick, and Smith (1998); Barker and Rayner (1998); Morris (1996); Barbieri and O'Hagan (1996); and Troughton and Godsill (1998), and applications of SSVS and related variants include McCulloch and Tsay (1994); Godsill and Rayner (1996); Godsill (1997); Godsill and Rayner (1998); Barnett, Kohn, and Sheather (1996); Troughton and Godsill (in press); Huerta and West (1999); Clyde, Desimone, and Parmigiani (1996).

1.3 PRINCIPAL CONTRIBUTIONS OF THIS ARTICLE

Section 2.1 presents the composite representation for model uncertainty problems. As in Carlin and Chib (1995), a product space is defined over all possible model parameters and a model indexing variable. Section 2.5 shows how to modify this framework to situations where it is more convenient to parameterize the space with some parameters shared between several models, as is the case in variable selection and nested models. In Carlin and Chib (1995) a Gibbs sampler is used to explore the product space. Subsequent sections of this article show that application of various alternative Gibbs or Metropolis schemes to the composite space can lead to the other well-known model samplers as special cases of the general framework. In particular, it is shown that reversible jump is obtained when a special form of Metropolis–Hastings proposal is applied; we note that a similar observation has independently been made by Besag (1997). These results add to the overall understanding of both reversible jump and composite space schemes and provide a review of the current ideas in the field, all considered within a single framework. The relationships developed shed some light upon the issues of pseudo-prior choice in the case of Carlin and Chib, and choice of proposal densities in the case of reversible jump. It is hoped that the general

framework may also lead to new classes of model space sampler that combine the benefits of several different schemes within the composite model. The final sections of the article give attention to devising efficient proposal distributions in cases where some analytic structure is present in the model. A simulation example is presented for the case of an autoregressive model with unknown model order, demonstrating the improvements achievable when model structure is taken into account.

2. RELATIONSHIPS BETWEEN DIFFERENT MODEL SPACE SAMPLERS

2.1 THE COMPOSITE REPRESENTATION FOR MODEL UNCERTAINTY PROBLEMS

We first define a composite model space for standard model selection problems in which no parameters are considered as "shared" between any two models. This is later modified to introduce more flexibility in shared parameter problems such as nested models and variable selection. The composite model is a straightforward modification of that used by Carlin and Chib (1995). Consider a "pool" of N parameters $\theta = (\theta_1, \ldots, \theta_N)$, such that θ_i has support Θ_i. The parameters θ_i may once again be vectors of differing dimensionality. In many applications the emphasis will be upon model classes in which the parameters do have variable dimension, especially when there is some structure or "nesting" of the models which relates the parameters of one model dimension to those of another. However, it is worth noting that all of the methods discussed here are equally applicable to cases where the candidate models have fixed dimensionality. A probability distribution is now defined over the entire product space of candidate models and their parameters; that is, $(k, \theta) \in \mathcal{K} \times \prod_{i=1}^{N} \Theta_i$, where $\mathcal{K}$ is the set of candidate model indices. The likelihood and prior structure are then defined in a corresponding way, as follows. For a particular k the likelihood depends only upon the corresponding parameter θ_k:

$$p(y|k, \theta) = p(y|k, \theta_k). \tag{2.1}$$

The model specification is completed by the parameter prior $p(\theta_k|k)$ and the model prior $p(k)$. The full posterior distribution for the composite model space can now be expressed as

$$p(k, \theta|y) = \frac{p(y|k, \theta_k)\, p(\theta_k|k)\, p(\theta_{-k}|\theta_k, k)\, p(k)}{p(y)}, \tag{2.2}$$

where θ_{-k} denotes the parameters *not* used by model k. All of the terms in this expression are defined explicitly by the chosen likelihood and prior structures except for $p(\theta_{-k}|\theta_k, k)$, the "prior" for the parameters in the composite model which are not used by model k. It is easily seen that any proper distribution can be assigned arbitrarily to these parameters without affecting the required marginals for the remaining parameters. We have given the general case, in which this prior can depend upon both k and the remaining model parameters. In many cases it will be convenient to assume that the unused parameters are a

priori independent of one another and also of θ_k. In this case we have that $p(\theta_{-k}|\theta_k, k) = p(\theta_{-k}|k) = \prod_{i \neq k} p(\theta_i|k)$ and the composite model posterior can be rewritten as

$$p(k, \theta|y) = \frac{p(y|k, \theta_k)\, p(\theta_k|k) \left(\prod_{i \neq k} p(\theta_i|k)\right) p(k)}{p(y)}. \tag{2.3}$$

This is the form of composite space used by Carlin and Chib (1995). The priors on the unused parameters θ_{-k} are referred to as "pseudo-priors" or linking densities in the Carlin and Chib model, appropriate choice of which is crucial to the effective operation of their algorithm. We will retain the general form as given in Equation (2.2), referring to the term $p(\theta_{-k}|\theta_k, k)$ as the pseudo-prior, although it should be noted that the simpler form of Equation (2.3) will often be used in practice, with consequent simplification of the posterior distribution.

The key feature of the composite model space is that the dimension remains fixed even when the model number k changes. This means that standard MCMC procedures, under the usual convergence conditions, can be applied to the problem of model uncertainty. For example, a straightforward Gibbs sampler applied to the composite model leads to Carlin and Chib's method, while we show later that a more sophisticated Metropolis–Hastings approach leads to reversible jump. In the following sections we show how to obtain these existing MCMC model space samplers as special cases of the composite space sampler.

2.2 CARLIN AND CHIB

The sampling algorithm of Carlin and Chib (1995) is easily obtained from the composite model by applying a Gibbs sampler to the individual parameters θ_i and to the model index k. The sampling steps, which may be performed in a random or deterministic scan, are as follows:

$$\begin{aligned} \theta_i &\sim p(\theta_i|\theta_{-i}, k, y) \propto \begin{cases} p(y|k, \theta_k)\, p(\theta_k|k), & i = k \\ p(\theta_i|\theta_{-i}, k), & i \neq k \end{cases} \\ k &\sim p(k|\theta, y) \propto p(y|k, \theta_k)\, p(\theta_k|k)\, p(\theta_{-k}|\theta_k, k)\, p(k). \end{aligned}$$

The method is rather impractical for problems with many candidate models since every parameter vector is sampled at each iteration, although Green and O'Hagan (1997) showed that this is in fact not necessary for strict convergence of the sampler. The problem can in fact be mitigated by replacing the Gibbs step for k with a Metropolis–Hastings step having the same target distribution; then it is necessary only to generate values for the parameters of any two models at each iteration rather than the complete step. Nevertheless, suitable choice of pseudo-priors is essential for efficient operation. Carlin and Chib suggested the use of pseudo-priors that are close to the posterior conditional for each model. We can see why this might be a good choice by analyzing the case when the pseudo-priors are set *exactly* to the posterior conditionals for each parameter and the individual parameters are assumed independent a priori

$$p(\theta_{-k}|\theta_k, k) = \prod_{i \neq k} p(\theta_i|y, k = i).$$

The sampling step for k is then found to reduce to

$$k \sim p(k|\theta, y) = p(k|y) = \int_{\Theta_k} p(\theta_k, k|y)\, d\theta_k.$$

In other words the model index sampling step becomes simply a draw from the true model posterior probability distribution $p(k|y)$ and does not depend upon the sampled parameter values θ_i. This is in some sense the ideal case since the aim of model uncertainty sampling is to design a sampler that explores model space according to $p(k|y)$. We can see then why choosing pseudo-priors that are close to the parameter conditionals is likely to lead to effective operation of the algorithm. Of course, the exact scheme is impractical for most models since $p(k|y)$ is typically unavailable in closed form, but this still gives some guidance as to what a suitable pseudo-prior might look like.

2.3 REVERSIBLE JUMP

The reversible jump sampler (Green 1995) achieves model space moves by Metropolis–Hastings proposals with an acceptance probability that is designed to preserve detailed balance within each move type. Suppose that we propose a move to model k' with parameters $\theta_{k'}$ from model k with parameters θ_k using a proposal distribution $q(k', \theta_{k'}; k, \theta_k)$. The acceptance probability in order to preserve detailed balance is given by

$$\alpha = \min\left(1, \frac{p(k', \theta_{k'}|y)\, q(k, \theta_k; k', \theta_{k'})}{p(k, \theta_k|y)\, q(k', \theta_{k'}; k, \theta_k)}\right). \tag{2.4}$$

This acceptance probability is expressed without use of measure-theoretic notation. Rather we have assumed that density functions exist with respect to, for example, Lebesgue measure for all of the distributions concerned, as will nearly always be the case in practice.

In implementation it will often be convenient to take advantage of any nested structure in the models or interrelationships between the parameters of different models in constructing effective proposal distributions, rather than proposing the entire new parameter vector as in (2.4). To take a very simple case, a fully nested model structure between models k and $k+1$ can easily be implemented by fixing the first k parameters in both models and making a proposal of the form $q(k+1, \theta_{k+1}; k, \theta_k) = q_1(k+1; k) q_2(\theta_{k+1}; \theta_k) = q_1(k+1; k) q_2(\theta_{k+1}^{(k+1)}|\theta_k)\delta_{\theta_k}(\theta_{k+1}^{(1:k)})$, where $\theta_{k+1}^{(1:k)}$ denotes the first k elements of θ_{k+1}. The reverse move is then of the form $q(k, \theta_k; k+1, \theta_{k+1}) = q_1(k; k+1)\delta_{\theta_{k+1}^{(1:k)}}(\theta_k)$ and the acceptance ratio simplifies to

$$\frac{p(k+1, \theta_{k+1}|y)\, q_1(k; k+1)}{p(k, \theta_k|y)\, q_1(k+1; k) q_2(\theta_{k+1}^{(k+1)}|\theta_k)}.$$

An example of the application of such a nested sampler, compared with a full parameter proposal of the form (2.4), is given in Section 2.7. More generally, relationships between parameters of different models can be used to good effect by drawing "dimension matching" variables u and u' from proposal distributions $q_2(u)$ and $q_2(u')$, and then forming $\theta_{k'}$ and θ_k as deterministic functions of the form $\theta_k = g(\theta_{k'}, u)$ and $\theta_{k'} = g(\theta_k, u')$. In this way it is straightforward to incorporate useful information from the current parameter vector θ_k

into the proposal for the new parameter vector $\theta_{k'}$. Provided that $\dim(\theta_{k'}, u) = \dim(\theta_k, u')$ (dimension matching), the acceptance probability is given by (Green 1995):

$$\alpha = \min\left(1, \frac{p(k', \theta_{k'}|y)\, q_1(k; k') q_2(u)}{p(k, \theta_k|y)\, q_1(k'; k) q_2(u')} \left|\frac{\partial(\theta_{k'}, u)}{\partial(\theta_k, u')}\right|\right)$$

which now includes a Jacobian term to account for the change of measure between (θ_k, u') and $(\theta_{k'}, u)$. Note that the basic form of reversible jump given above in Equation (2.4) is obtained from this formula when we set $\theta_k = g(\theta_{k'}, u) = u$ and $\theta_{k'} = g(\theta_k, u') = u'$, so that the Jacobian term is unity.

It is worth commenting that the earlier jump diffusion methods for model uncertainty (Grenander and Miller 1991; Grenander and Miller 1994; Phillips and Smith 1994) can be considered as a special version of the reversible jump scheme in which model jumps are proposed with exponentially distributed time gaps and parameter moves are performed using discretised Langevin diffusions. Hence we do not address these methods further here.

2.3.1 Reversible Jump Derived From the Composite Model

We now show that Green's reversible jump sampler can be obtained by applying a special form of Metropolis–Hastings (M–H) proposal to the composite model space. We derive the general form given in (2.4), noting as above that nested and other forms can be obtained from this general case provided that dimension matching constraints are carefully incorporated.

Consider a proposal from the current state of the composite model (k, θ) to a new state (k', θ') that takes the form:

$$q(k', \theta'; k, \theta) = q_1(k'; k)\, q_2(\theta'_{k'}; \theta_k)\, p(\theta'_{-k'}|\theta'_{k'}, k').$$

This proposal, which forms a joint distribution over all elements of k and θ, is split into three component parts: the model index component $q_1(k'; k)$, which proposes a move to a new model index, k'; a proposal for the parameters used by model k', $q_2(\theta'_{k'}; \theta_k)$; and a proposal for the remaining unused parameters which is chosen to equal the pseudo-prior $p(\theta'_{-k'}|\theta'_{k'}, k')$. We thus have a joint proposal across the whole state space of parameters and model index that satisfies the Markov requirement of the M–H method as it depends only upon the current state (k, θ) to make the joint proposal (k', θ'). There are now no concerns about a parameter space with variable dimension since the composite model retains constant dimensionality whatever the value of k and any issues of convergence can be addressed by reference to standard M–H results in the composite space.

The acceptance probability for this special form of proposal is given, using the standard M–H procedure, by

$$\begin{aligned}
\alpha &= \min\left(1, \frac{q(k, \theta; k', \theta')\, p(k', \theta'|y)}{q(k', \theta'; k, \theta)\, p(k, \theta|y)}\right) \\
&= \min\left(1, \frac{q_1(k; k')\, q_2(\theta_k; \theta'_{k'})\, p(\theta_{-k}|\theta_k, k)\, p(k', \theta'_{k'}|y)\, p(\theta'_{-k'}|\theta'_{k'}, k')}{q_1(k'; k)\, q_2(\theta'_{k'}; \theta_k)\, p(\theta'_{-k'}|\theta'_{k'}, k')\, p(k, \theta_k|y)\, p(\theta_{-k}|\theta_k, k)}\right) \\
&= \min\left(1, \frac{q_1(k; k')\, q_2(\theta_k; \theta'_{k'})\, p(k', \theta'_{k'}|y)}{q_1(k'; k)\, q_2(\theta'_{k'}; \theta_k)\, p(k, \theta_k|y)}\right).
\end{aligned}$$

This last line is exactly the acceptance probability for the basic reversible jump sampler (2.4) with the proposal distribution factored in an obvious way into two components $q_1(.)$ and $q_2(.)$. We see that the acceptance probability is independent of the value of any parameters which are unused by both models k and k' (their pseudo-priors cancel in the acceptance probability); nor are their values required for generating a proposal at the next iteration. Hence the sampling of these is a "conceptual" step only which need not be performed in practice. This feature is a strong point of the reversible jump method compared with the Gibbs sampling version of the Carlin and Chib method, which requires samples for all parameters, including pseudo-parameters, at every iteration. Conversely, it is a very challenging problem to construct effective proposal distributions for reversible jump methods in complex modeling scenarios, especially in cases where there is no obvious nested structure to the models or other interrelationships between the parameters of the different models; in these cases the Carlin and Chib method, which allows blocking of the parameters within a single model in a way that is not possible for reversible jump, may have the advantage. It is interesting, however, to see that both schemes can be derived as special cases of the composite space sampler.

Convergence properties of the reversible jump scheme derived in the special way given here can now be inherited directly from the Metropolis–Hastings algorithm operating on the fixed dimesion composite space. Specifically, irreducibility and aperiodicity of the composite space sampler will ensure the convergence of the chain to the target distribution and the validity of ergodic averages (Robert and Casella 1999).

In independent work on variable selection methods by Dellaportas, Forster, and Ntzoufras (1997), it is observed that the composite space sampler can be obtained, for the two-model case, taking reversible jump as the starting point. This interesting result is related to ours. We believe, however, that our work goes beyond this by showing that reversible jump may be obtained purely from fixed-dimensional considerations on the composite space, and hence that convergence properties are inherited directly from the fixed-dimension M–H method. Our derivation of reversible jump is also very straightforward, requiring no measure-theoretic detail beyond that associated with a standard M–H sampler.

2.3.2 Proposing From Full Posterior Conditionals and MC[3]

In a similar vein to the suggestions made above for the Carlin and Chib method, a possible version of reversible jump would use the full posterior conditional $p(\theta_{k'}|k', y)$ as the proposal density $q_2(.)$ in the above description. We can then employ the identity $\frac{p(k,\theta|y)}{p(\theta|k,y)} = p(k|y)$ [Besag (1989) used this basic identity to find prediction densities; Chib (1995) and Chib and Greenberg (1998) used a related identity to calculate Bayes factors, for example] to obtain the following acceptance probability:

$$\alpha = \min\left(1, \frac{p(k'|y)q_1(k;k')}{p(k|y)q_1(k';k)}\right).$$

This can be recognized as the acceptance probability of a standard Metropolis–Hastings method with the posterior model probability $p(k|y)$ as the target distribution and using proposals $q_1(k'|k)$ for the model moves. Note that the acceptance probability is independent of parameter values, depending only upon the proposal distribution for model order

and the posterior odds ratio $p(k'|y)/p(k|y)$. Inference about parameter values can then be made conditional upon the current model index k using standard MCMC. Such a scheme has been used for decomposable graphical models in the MC^3 method of Madigan and York (1995). Stark, Fitzgerald, and Hladky (1997) suggested a similar scheme for use with changepoint models. They pointed out that the parameters generated in proposing from the full conditional distribution $p(\theta_{k'}|k', y)$ can be used in a subsequent Gibbs sampling step for $\theta_{k'}$ if the move to model k' is accepted.

In the (relatively rare) cases where $p(k|y)$ or equivalently the value of the full conditional $p(\theta_k|k, y)$ at all values of θ_k is available analytically, use of conditional parameter distributions as reversible jump proposals would lead to excellent exploration of model space. This would suggest that reversible jump proposals should be designed to approximate as closely as possible the parameter conditionals in order to come close to the performance of the scheme when parameter conditionals are readily available in exact form. This is similar in principle to Carlin and Chib's suggestion that pseudo-priors be chosen close to the parameter conditionals in their method.

2.4 USE OF PARTIAL ANALYTIC STRUCTURE IN REVERSIBLE JUMP PROPOSALS

The exact scheme of the last section is, of course, not available for most models of practical interest. Nevertheless, many useful models will have what we term *partial analytic* structure; that is, we have the full conditional in closed form for some subvector of $\theta_{k'}$, the vector of parameters which are used by the new model k'; in other words $p((\theta_{k'})_{\mathcal{U}}|(\theta_{k'})_{-\mathcal{U}}, k', y)$ is available for some subset of the parameters, indexed by a set $\mathcal{U}$. If we suppose that an equivalent subset of parameters $(\theta_k)_{-\mathcal{U}}$, with the same dimensionality as $(\theta_{k'})_{-\mathcal{U}}$, is present in the current model k, we might choose a reversible jump proposal distribution which sets $(\theta_{k'})_{-\mathcal{U}} = (\theta_k)_{-\mathcal{U}}$ and proposes the remaining parameter vector $(\theta_{k'})_{-\mathcal{U}}$ from its full conditional, $p((\theta_{k'})_{\mathcal{U}}|(\theta_{k'})_{-\mathcal{U}}, k', y)$. The reverse move would set $(\theta_k)_{-\mathcal{U}} = (\theta_{k'})_{-\mathcal{U}}$ and propose the remaining parameters in model k from their conditional $p((\theta_k)_{\mathcal{U}}|(\theta_k)_{-\mathcal{U}}, k', y)$. (Note that in general $(\theta_k)_{\mathcal{U}}$ and $(\theta_{k'})_{\mathcal{U}}$ will be of differing dimensionality.) The reversible jump acceptance probability for such a move can then be derived as

$$\alpha = \min\left(1, \frac{p\big(k'|(\theta_{k'})_{-\mathcal{U}} = (\theta_k)_{-\mathcal{U}}, y\big)\, q_1(k; k')}{p(k|(\theta_k)_{-\mathcal{U}}, y)\, q_1(k'; k)}\right), \tag{2.5}$$

where $p(k'|(\theta_{k'})_{-\mathcal{U}}, y) = \int_{(\Theta_k)_{\mathcal{U}}} p(k', (\theta_{k'})_{\mathcal{U}}|(\theta_{k'})_{-\mathcal{U}}, y)\, d(\theta_{k'})_{\mathcal{U}}$. A typical example where this might be used is the linear Gaussian model with conjugate priors, where the full conditional for the linear parameters is available. $(\theta_{k'})_{\mathcal{U}}$ might then be chosen to be the linear parameters for model k', while $(\theta_{k'})_{-\mathcal{U}}$ could be the remaining unknown prior hyperparameters—such as noise variances—which are common to both models k and k'. These parameters, being of fixed dimensionality, can then be sampled in a separate step using a standard fixed-dimension MCMC method such as the Gibbs sampler or Metropolis–Hastings. Of course, a more sophisticated scheme might also include a random proposal to change the value of these "core" parameters within the reversible jump proposal. In this way we can also deal with the case where the set of core parameters $(\theta_k)_{-\mathcal{U}}$ depends on k

and hence the dimensionality of $(\theta_k)_{-\mathcal{U}}$ may also vary with k.

Note once again that the acceptance probability does not depend upon the sampled parameter values $(\theta_{k'})_{\mathcal{U}}$ or $(\theta_k)_{\mathcal{U}}$. In this case it depends upon the model proposal distributions and the posterior odds *conditional upon* $(\theta_{k'})_{-\mathcal{U}} = (\theta_k)_{-\mathcal{U}}$. In cases where $(\theta_k)_{-\mathcal{U}}$ can be given a similar interpretation in both models (the parameters are "common" to models k and k') this scheme is likely to yield a simple and effective model space sampler which takes advantage of any analytic structure within the model. A simulation example of this approach is given in Section 2.7.

2.5 A COMPOSITE SPACE FOR NESTED MODELS AND VARIABLE SELECTION

Thus far we have considered only basic model selection problems in which parameters are not shared between different models, as will be the case for nested models and variable selection problems. Although it is always possible to represent such models within the basic framework described above, it will be convenient to define a slight generalization that specifically accounts for parameter "overlap" between models. As before, we have a "pool" of N parameters $\theta = (\theta_1, \ldots, \theta_N)$. We now, however, introduce a set of indices for each model which indicates which of the θ_i's are used by model k. Specifically, $\mathcal{I}(k) = \{i_1(k), i_2(k), \ldots i_{l(k)}(k)\} \subseteq \{1, 2, \ldots, N\}$ defines the model-dependency of the likelihood as follows:

$$p(y|k, \theta) = p(y|k, \theta_{\mathcal{I}(k)}), \tag{2.6}$$

where $\theta_{\mathcal{I}(k)} = (\theta_i;\, i \in \mathcal{I}(k))$ denotes the parameters used by model k. Much as before, the full composite posterior can be expressed as:

$$p(k, \theta|y) = \frac{p(y|k, \theta_{\mathcal{I}(k)})\, p(\theta_{\mathcal{I}(k)}|k)\, p(\theta_{-\mathcal{I}(k)}|\theta_{\mathcal{I}(k)}, k)\, p(k)}{p(y)}, \tag{2.7}$$

where $\theta_{-\mathcal{I}(k)} = (\theta_i;\, i \in \{1, \ldots, N\} - \mathcal{I}(k))$ denotes the parameters *not* used by model k. The term $p(\theta_{-\mathcal{I}(k)}|\theta_{\mathcal{I}(k)}, k)$ is once again a "pseudo-prior" for which any proper distribution will suffice.

Using this framework, some convenient (but not unique) parameterizations of the composite space are:

- *Standard model selection.* In the basic model selection problem we associate one parameter vector with each model, so we can simply use $k \in \{1, \ldots, N\}$ and $\mathcal{I}(k) = \{k\}$. The model is assumed not to be nested, so no elements of different parameter vectors are considered to be common across different models. Of course, all model uncertainty problems can be formulated in this way, but it will often not be convenient to use this representation for computational reasons.

- *Nested models.* In nested models it is assumed that parameters from the model of order k have the same interpretation as the first k parameters in the model of order $k + 1$. In this case we have $k \in \{1, \ldots, N\}$, as before, but now $\mathcal{I}(k) = \{1, \ldots, k\}$.

• *Variable selection.* In variable selection problems the model conditional likelihood can depend upon any combination of the available parameters. In this case a natural parameterization for k is as a binary N-vector; that is, $k = [k_1, k_2, \ldots, k_N] \in \{0, 1\}^N$, and $\mathcal{I}(k) = \{i : k_i = 1\}$. Each element of k then "switches" a particular dependent variable in or out of the model (e.g., setting $k = [0, \ldots, 0]$ corresponds to the case where the data depend upon none of the candidate variables). This is a pure variable selection problem in which the model conditioned likelihood is truly independent of all those candidate variables which have $k_i = 0$. Many problems that involve latent indicator variables can be viewed as variable selection problems, and we thus use the term here in its most general sense to include all of these variants on the problem. This parameterization of the variable selection problem is equivalent to that used by Kuo and Mallick (1998) and Geweke (1996). Note that the stochastic search variable selection (SSVS) methods of George and McCulloch (1993) are not quite the same as this since the likelihood in their case is of fixed form for all models, depending upon all parameters within every model. Model uncertainty is then built in through a prior structure which enforces very small values for those parameters which are switched "off" in the model. This avoids some of the difficulties of working with a variable dimension parameter space. Within the framework of the composite model we could achieve this configuration by setting $\mathcal{I}(k) = \{1, \ldots N\} \, \forall k$. The likelihood (2.6) is then taken as independent of k and distinction between different k is achieved purely through the prior distributions on the θ_i's and k. We will consider here the first formulation in which parameters can be switched out of the model completely; that is, the likelihood is completely independent of θ_i when $k_i = 0$. Methods based upon these principles find application not only in traditional "variable selection" problems but in many other areas where individual effects can be modeled via latent indicator variables (see references in Section 1.2).

The results of the previous sections applied to the standard model selection problem are readily adapted to the more general framework of this section. We now go on to discuss MCMC variable selection within the composite model space framework.

2.6 MCMC VARIABLE SELECTION

Using the parameterization described earlier in which k is a binary vector of parameter "indicators," MCMC variable selection methods are obtained immediately by the application of a Gibbs sampler to the parameter space partitioned as $(k_1, k_2, \ldots, k_N, \theta_1, \theta_2, \ldots, \theta_N)$. If, for simplicity, we omit any additional hyperparameters such as noise variances, which are often considered to be common to all models, then the following sampling scheme is obtained, which is essentially the same as that of Kuo and Mallick (1998):

$$\begin{aligned} \theta_i &\sim p(\theta_i|\theta_{-i}, k, y) \propto \begin{cases} p(y|\theta_{\mathcal{I}(k)}, k)\, p(\theta_{\mathcal{I}(k)}|k)\, p(\theta_{-\mathcal{I}(k)}|\theta_{\mathcal{I}(k)}, k), & k_i = 1 \\ p(\theta_i|\theta_{-i}, k), & k_i = 0 \end{cases} \\ k_i &\sim p(k_i|k_{-i}, \theta, y) \propto p(y|\theta_{\mathcal{I}(k)}, k)\, p(\theta_{\mathcal{I}(k)}|k)\, p(\theta_{-\mathcal{I}(k)}|\theta_{\mathcal{I}(k)}, k). \end{aligned}$$

The individual parameters θ_i are thus sampled either from their posterior conditional or from their pseudo-prior, depending upon the value of k_i. Clearly schemes which use other

types of MCMC in the moves or choose alternative blocking strategies to yield improved performance can also be devised (see, e.g., Godsill and Rayner 1996, 1997; Barnett, Kohn, and Sheather 1996; Carter and Kohn 1996; Troughton and Godsill in press).

The fact that the pseudo-priors can be chosen arbitrarily in exactly the same way as for the standard model selection problem is not often noted within a variable selection framework. One practically useful example of this fact, in some variable selection models, such as those involving linear conditionally Gaussian assumptions for the parameters, is to choose the pseudo-prior for each parameter θ_i to be the conditional posterior for θ_i with $k_i = 1$; that is, set

$$p(\theta_i|k_i = 0, \theta_{-i}, k_{-i}) = p(\theta_i|k_i = 1, \theta_{-i}, k_{-i}, y).$$

In the basic Gibbs sampling framework summarized above, the sampling step for k_i then reduces to:

$$k_i \sim p(k_i|\theta_i, \theta_{-i}, k_{-i}, y) = p(k_i|\theta_{-i}, k_{-i}, y) = \int_{\theta_i} p(k_i, \theta_i|\theta_{-i}, k_{-i}, y) d\theta_i.$$

When associated with the conditional draw of θ_i from its conditional posterior $p(\theta_i|\theta_{-i}, k, y)$ we see that the approach is equivalent to a blocking scheme which draws jointly for (θ_i, k_i) using the decomposition $p(\theta_i, k_i|\theta_{-i}, k_{-i}, y) = p(\theta_i|k, y)p(k_i|\theta_{-i}, k_{-i}, y)$. Such blocking schemes have been found empirically to give much improved performance over straightforward single-move Gibbs samplers both in outlier analysis (Godsill and Rayner 1996, 1998; Godsill 1997; Barnett, Kohn, and Sheather 1996) and variable selection for nonlinear time series (Troughton and Godsill in press). This blocking procedure can also be viewed as equivalent to that used by Geweke (1996), who reparameterized the problem with δ-functions in the prior for variables which are not used in the model. In these cases the integral required can easily be performed analytically. In other cases, improved performance could be achieved over the "single move" Gibbs Sampler by setting the pseudo-priors to some suitable approximation to the conditional posterior in a similar fashion to Carlin and Chib's proposal for the basic model selection problem.

2.7 EXAMPLE

To illustrate the principle of using partial analytic structure in reversible jump proposals we examine a simple time series autoregression model uncertainty problem:

$$x_t = \sum_{i=1}^{k} a_i^{(k)} x_{t-i} + e_t, \qquad e_t \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_e^2),$$

where $a^{(k)} = (a_i^{(k)}; i = 1, \ldots, k)$ are the AR coefficients for a model of order k. For simplicity we side-step issues of stationarity and work with the conditional likelihood, which approximates the exact likelihood well for large N (Box, Jenkins, and Reinsel 1994):

$$p(x|a^{(k)}, \sigma_e^2, k) = \prod_{i=1}^{N} \mathrm{N}\left(x_t - \sum_{i=1}^{P} a_i^{(k)} x_{t-i}\right), \qquad x = [x_1 \ \ldots \ x_N].$$

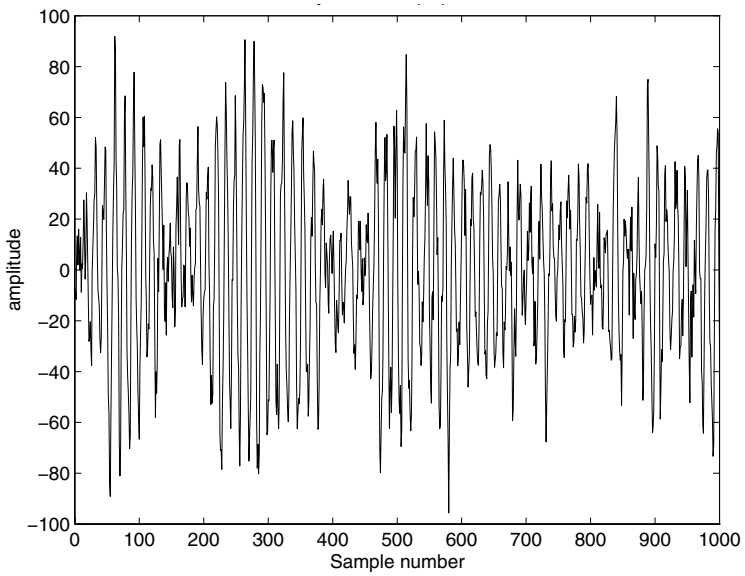

Figure 1. Synthetic AR(10) data.

Conjugate normal and inverted Gamma priors are assumed for $a^{(k)}$ and σ_e^2. A uniform prior is assumed for k over a range $1, \ldots, k_{\max}$, where $k_{\max}$ was set at 30 in this example. Some partial analytic structure is then available in the form of the conditional distribution for $a^{(k)}$, $p(a^{(k)}|x, \sigma_e^2, k)$, which is multivariate Gaussian (Box, Jenkins, and Reinsel 1994). Thus, we set $\theta_k = (a^{(k)}, \sigma_e^2)$, $(\theta_k)_{\mathcal{U}} = a^{(k)}$, and $(\theta_k)_{-\mathcal{U}} = \sigma_e^2$. The acceptance probability for model moves, following (2.5), is then:

$$\alpha = \min\left(1, \frac{p(k'|\sigma_e^2, X)q(k; k')}{p(k|\sigma_e^2, x)q(k'; k)}\right),$$

where $p(k|\sigma_e^2, x) = \int_{a^{(k)}} p(a^{(k)}, k|\sigma_e^2, x)\, da^{(k)}$, which is obtained analytically. σ_e^2 is updated at each iteration using a standard Gibbs sampling step.

One thousand data points are simulated from an order 10 model with coefficients $a^{(10)} = [0.9402, -0.4300, 0.4167, -0.4969, 0.4771, -0.5010, 0.0509, -0.2357, 0.4024, -0.1549]$, and $\sigma_e^2 = 100$, as shown in Figure 1. We chose a relatively large dataset to ensure that the likelihood expression is accurate and also because this highlighted the differences between the two schemes considered. The schemes were: the method above based upon the partial analytic structure of the model, and a simple reversible jump implementation which proposes new parameters from an iid Gaussian; that is,

$$\theta_{k'} = \begin{array}{ll} [\sigma_e^2, a_1^{(k)} \ \ldots \ a_{k'}^{(k)}], & k' \le k, \\ \left[\theta_k \, u_1 \ldots u_{k'-k}\right], \quad u_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma_u^2), & k' > k \end{array}.$$

The acceptance probability for such a proposal is (see Green 1995, eq. (8)), for $k' > k$:

$$\alpha = \min\left(1, \frac{p(a^{(k')}, k'|\sigma_e^2, x)q(k; k')}{p(a^{(k)}, k|\sigma_e^2, x)q(k'; k)\prod_{j=k+1}^{k'} \mathrm{N}(\theta_j|0, \sigma_u^2)}\right)$$

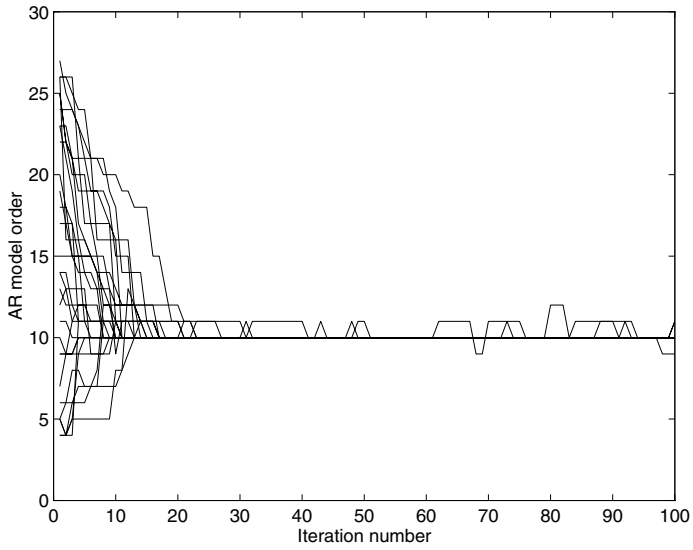

Figure 2. Model order evolution using the partial analytic sampler.

and the form of the fraction term is inverted for $k' < k$. We refer to this simple reversible jump implementation as the "stepwise" sampler. In all other respects the two methods are identical, both including a "within model" Gibbs move for $a^{(k)}$ and σ_e^2 at each iteration. The prior for $a^{(k)}$ was $\mathrm{N}_k(0, 0.1\ I)$, and for σ_e^2, $\mathrm{IG}(10^{-5}, 10^{-5})$. The proposal distribution for the model orders, $q(.;.)$ has a discretised Laplacian shape, centered on the current model order. Note that this allows fairly frequent proposals to model orders which are distant from the current model.

The initial model order was assigned randomly from a uniform distribution over integers 1 to 30. The AR parameters were initialized to zero. The first step of the sampler was a Gibbs draw for σ_e^2, so this does not require initialization. Results for 30 runs of the partial analytic sampler are superimposed in Figures 2 and 3, showing the consequences of randomly assigned initial model orders. We show only the initial hundred iterations as the sampler has always stabilized within the first few tens of iterations. By contrast we show the same results for the stepwise sampler under exactly the same initialization conditions and proposing the new parameters from a zero mean normal distribution with variance $\sigma_u^2 = 0.1$. Note the different axis scaling for iteration number. The stepwise sampler has often not settled down within the first 1,000 iterations and changes state relatively rarely. We do not claim that to have optimized the standard reversible jump implementation here as there are many possible options; however, this comparison gives a reasonable flavor of the improvements which are achievable automatically, without any parameter tuning, simply through the use of the analytic structure of the model.

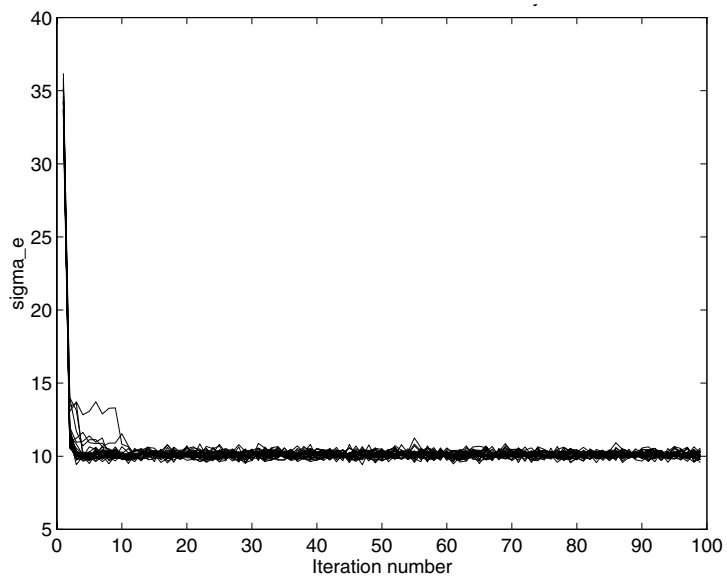

Figure 3. Evolution of σ_e using the partial analytic sampler.

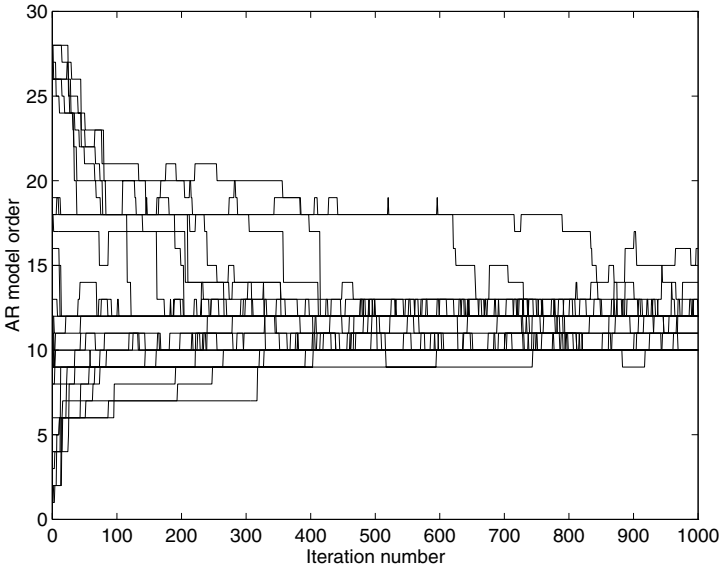

Figure 4. Model order evolution using the stepwise sampler.

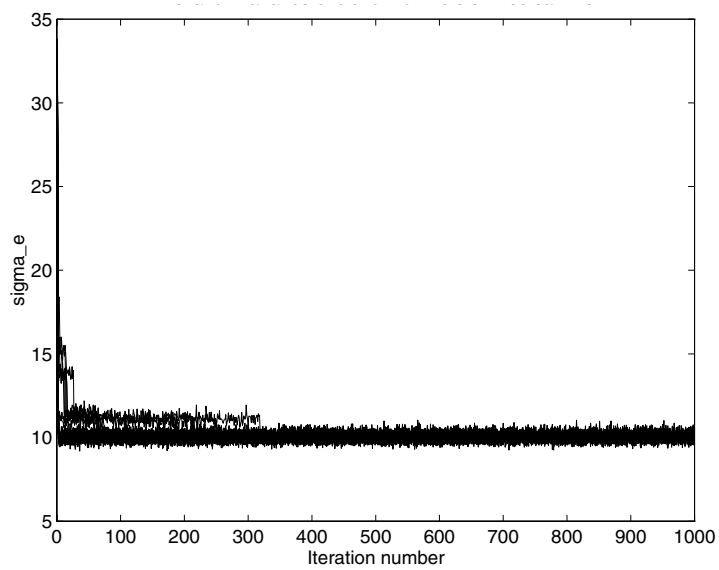

Figure 5. Evolution of σ_e using the stepwise sampler.

3. DISCUSSION

This article presents a unifying framework for MCMC model uncertainty schemes, which includes as special cases the existing methods for treatment of the problem using reversible jump or the Carlin and Chib approach. We have demonstrated further that there are close relationships between these methods and MCMC variable selection techniques for model space sampling. Simple analysis has shown that pseudo-priors (in the case of Carlin and Chib and MCMC variable selection) and parameter proposal distributions (in the case of reversible jump) which are designed to be close to the full posterior conditional for the parameters are likely to lead to very effective performance of both methods. Furthermore, we have proposed methods for taking advantage of partial analytic structure in a particular model to achieve efficient model space moves.

The reversible jump scheme is a very effective way to apply a Metropolis–Hastings sampler to the composite space. One of its major advantages over Carlin and Chib's approach is that the values of parameters from models other than the two being compared at any given iteration (i.e., k and k') need not be updated or stored and pesudo-priors need not be devised. This is crucial for the large (or even infinite) sets of models which might need to be considered. However, it should be noted that many problems exist where there is no obvious way to construct a reversible jump proposal to a new model based on the parameters of the current model in the MCMC output. In such cases it would be desirable to use some elements of parameter blocking for the new model's parameters. Such a scheme is feasible for the Carlin and Chib approach and its Metropolized versions, although choice of pseudo-prior will still be a troublesome aspect. It would seem that some hybrid approach that adopts the

best aspects of reversible jump and Carlin and Chib is required, although the development of such a procedure is still an open issue. The composite framework presented here, however, provides a possible starting point for such a scheme.

We believe that the solution of model uncertainty problems using MCMC is an important field in which there have been significant advances over recent years. However, there are a number of technical challenges still to be solved. As with any MCMC procedure, convergence assessment will be an important aspect of any practical scheme. This is a topic that is still in its infancy for the variable dimension samplers, although some suggestions and insight can be found in Brooks and Giudici (1999), who proposed monitoring the statistics of multiple independent chains in the same spirit as Gelman and Rubin (1992). Our practical experience confirms that a parallel chain approach, initializing samplers in models with widely differing characteristics, gives important diagnostic information. There are, of course, modeling scenarios where the current methods are infeasible. As implied elsewhere in this article, this may occur when there is little or no structural relationship between different candidate models and hence it is very hard to construct effective model jumping proposals. If in addition the models concerned are individually complex and require sophisticated blocking strategies to sample from, even in the single model case, then current methods are likely to be ineffective in the presence of model uncertainty. In these cases the only option currently will be to perform direct estimation of marginal likelihoods within each model, a procedure that is not readily implemented when the number of candidate models is very large. It is these cases then that demand the attention of researchers over coming years and we hope that the material presented here provides one possible framework for the development of improved strategies.

ACKNOWLEDGMENTS

Thanks are due to the anonymous reviewers for their helpful and insightful comments. Also many thanks to Peter Green and Sid Chib for many useful discussions and suggestions on various aspects of the work. Finally, thanks to Peter Green and John Forster for organizing and inviting me to the 1997 HSSS Model Selection workshop in the New Forest at which I first presented this work.

[Received TKKK. Revised TKKK.]

REFERENCES

Barbieri, M., and O'Hagan, A. (1996), "A Reversible Jump MCMC Sampler for Bayesian Analysis of ARMA Time Series," Technical Report, Università Roma: La Sapienza.

Barker, S., and Rayner, P. J. W. (1998), "Unsupervised Image Segmentation," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing* (vol. 5), pp. 2757–2760.

Barnett, G., Kohn, R., and Sheather, S. (1996), "Bayesian Estimation of an Autoregressive Model Using Markov Chain Monte Carlo," *Journal of Econometrics*, 74, 237–254.

Besag, J. (1989), "A Candidate's Formula—A Curious Result in Bayesian Prediction," *Biometrika*, 76, 183.

——— (1997), Discussion of "Bayesian Analysis of Mixtures With an Unknown Number of Components," by S. Richardson and P. Green, *Journal of the Royal Statistical Society*, Series B, 59, 774.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994), *Time Series Analysis, Forecasting and Control* (3rd ed.), Englewood Cliffs, NJ: Prentice Hall.

Brooks, S. P., and Giudici, P. (1999), "Convergence Assessment for Reversible Jump MCMC Simulations," in *Bayesian Statistics 6*, eds. J. Bernardo, J. Berger, A. Dawid, and A. Smith, Oxford: Oxford University Press, pp. 733–742.

Carlin, B. P., and Chib, S. (1995), "Bayesian Model Choice via Markov Chain Monte Carlo," *Journal of the Royal Statistical Society*, Series B, 57, 473–484.

Carter, C., and Kohn, R. (1996), "Markov Chain Monte Carlo in Conditionally Gaussian State Space Models," *Biometrika*, 83, 589–601.

Chib, S. (1995), "Marginal Likelihood From the Gibbs Output," *Journal of American Statistical Association*, 90, 1313–1321.

Chib, S., and Greenberg, E. (1998), "Analysis of Multivariate Probit Models," *Biometrika*, 85, 347–361.

Clyde, M., Desimone, H., and Parmigiani, G. (1996), "Prediction via Orthogonalized Model Mixing," *Journal of American Statistical Association*, 91, 1197–1208.

Dellaportas, P., Forster, J., and Ntzoufras, I. (1997), "On Bayesian Model and Variable Selection Using MCMC," Technical Report, Department of Statistics, Athens University of Economy and Business.

Denison, D., Mallick, B., and Smith, A. (1998), "Automatic Bayesian Curve Fitting," *Journal of the Royal Statistical Society*, Series B, 60, 333–350.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Gelman, A., and Rubin, D. (1992), "Inference From Iterative Simulations Using Multiple Sequences," *Statistical Science*, 7, 457–511.

Geman, S., and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

George, E. I., and McCulloch, R. E. (1993), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881–889.

——— (1996), "Approaches for Bayesian Variable Selection," *Statistica Sinica*, 7, 339–374.

Geweke, J. (1996), "Variable Selection and Model Comparison in Regression," in *Bayesian Statistics V*, eds. J. Bernardo, J. Berger, A. Dawid, and A. Smith, Oxford: Oxford University Press, pp. 609–620.

Godsill, S. J. (1997), "Bayesian Enhancement of Speech and Audio Signals Which can be Modelled as ARMA Processes," *International Statistical Review*, 65, 1–21.

Godsill, S. J., and Rayner, P. J. W. (1996), "Robust Treatment of Impulsive Noise in Speech and Audio Signals," in *Bayesian Robustness—Proceedings of the Workshop on Bayesian Robustness* (vol. 29), eds. J. Berger, B. Betro, E. Moreno, L. Pericchi, F. Ruggeri, G. Salinetti, and L. Wasserman, Hayward, CA: IMS, pp. 331–342.

——— (1998), "Robust Reconstruction and Analysis of Autoregressive Signals in Impulsive Noise Using the Gibbs Sampler," *IEEE Transactions on Speech and Audio Processing*, 6, 352–372.

Green, P., and O'Hagan, A. (1997), "Carlin and Chib do not Need to Sample From Pseudo-Priors," unpublished report.

Green, P. J. (1995), "Reversible Jump Markov-Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732.

Grenander, U., and Miller, M. I. (1991), "Jump-Diffusion Processes for Abduction and Recognition of Biological Shapes," Technical Report, Electronic Signals and Systems Research Laboratory, Washington University.

——— (1994), "Representations of Knowledge in Complex Systems," *Journal of the Royal Statistical Society*, Series B, 56, 549–603.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57, 97–109.

Hoeting, J., Raftery, A., and Madigan, D. (1996), "A Method for Simultaneous Variable Selection and Outlier Identification in Linear-Regression," *Computational Statistics and Data Analysis*, 22, 251–270.

Huerta, G., and West, M. (1999), "Priors and Component Structures in Autoregressive Time Series Models," *Journal of the Royal Statistical Society*, Series B, 61, 881–899.

Kuo, L., and Mallick, B. (1998), "Variable Selection for Regression Models," *Sankhya*, Series B, 60, 65–81.

Madigan, D., and York, J. (1995), "Bayesian Graphical Models for Discrete Data," *International Statistical Review*, 63, 215–232.

McCulloch, R. E., and Tsay, R. S. (1994), "Bayesian Analysis of Autoregressive Time Series via the Gibbs Sampler," *Journal of Time Series Analysis*, 15, 235–250.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21, 1087–1091.

Morris, R. (1996), "A Sampling Based Approach to Line Scratch Removal From Motion Picture Frames," in *Proceedings IEEE International Conference on Image Processing*, Lausanne, Switzerland.

Phillips, D. B., and Smith, A. F. M. (1994), "Bayesian Model Comparison via Jump Diffusions," Technical Report TR-94-20, Imperial College.

Raftery, A. (1996), "Hypothesis Testing and Model Selection," in *Markov Chain Monte Carlo in Practice*, eds. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, London: Chapman and Hall, pp. 163–187.

Raftery, A., Madigan, D., and Hoeting, J. (1997), "Bayesian Model Averaging for Linear Regression Models," *Journal of the American Statistical Association*, 92, 179–191.

Richardson, S., and Green, P. (1997), "Bayesian Analysis of Mixtures With an Unknown Number of Components," *Journal of the Royal Statistical Society*, Series B, 59, 731–758.

Robert, C., and Casella, G. (1999), *Monte Carlo Statistical Methods*, New York: Springer Verlag.

Stark, J., Fitzgerald, W., and Hladky, S. (1997), "Multiple-Order Markov Chain Monte Carlo Sampling Methods with Application to a Changepoint Model," Technical Report CUED/F-INFENG/TR.302, Department of Engineering, University of Cambridge.

Troughton, P., and Godsill, S. J. (1998), "A Reversible Jump Sampler for Autoregressive Time Series," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (vol. IV), pp. 2257–2260.

——— (in press), "MCMC Methods for Restoration of Nonlinearly Distorted Autoregressive Signals," *Signal Processing*, 81.

West, M., and Harrison, J. (1997), *Bayesian Forecasting and Dynamic Models* (2nd ed.), New York: Springer-Verlag.