Transdimensional Markov Chains: A Decade of Progress and Future Perspectives

Scott A. SISSON

The last 10 years have witnessed the development of sampling frameworks that permit the construction of Markov chains that simultaneously traverse both parameter and model space. Substantial methodological progress has been made during this period. In this article we present a survey of the current state of the art and evaluate some of the most recent advances in this field. We also discuss future research perspectives in the context of the drive to develop sampling mechanisms with high degrees of both efficiency and automation.

KEY WORDS: Automatic sampling; Convergence assessment; Efficient chains; Reference priors; Reversible jump; Software.

1. INTRODUCTION

Simultaneous inference on both model and parameter space is an issue that is fundamental to modern statistical practice. In general, for observed data $\mathbf{x}$ we might consider a countable set of models, $\mathcal{M} = \{M_1, M_2, \dots\}$, indexed by a parameter, $k \in \mathcal{K}$, each with a parameter vector defined on $\boldsymbol{\theta}_k \in \Theta_k$ of length n_k. Under a Bayesian framework, we would relate each model to a posterior distribution,

$$M_k: \quad \tilde{\pi}_k(\boldsymbol{\theta}_k|\mathbf{x}) = \frac{\pi_k(\boldsymbol{\theta}_k|\mathbf{x})}{m_k(\mathbf{x})} \propto L_k(\mathbf{x}|\boldsymbol{\theta}_k)p_k(\boldsymbol{\theta}_k),$$

generally known only up to a constant of proportionality, $m_k(\mathbf{x})^{-1}$, where L_k and p_k denote the likelihood and parameter prior under model M_k. Explicitly expressing $m_k(\mathbf{x}) = \int_{\Theta_k} L_k(\mathbf{x}|\boldsymbol{\theta}_k)p_k(\boldsymbol{\theta}_k)\,d\boldsymbol{\theta}_k$ as the marginal or predictive densities of $\mathbf{x}$ under model M_k, the normalized posterior probability of model M_k is given by

$$M_k(\mathbf{x}) = \frac{\rho_k m_k(\mathbf{x})}{\sum_{i=1}^{|\mathcal{K}|} \rho_i m_i(\mathbf{x})} = \left(1 + \sum_{i \neq k} \frac{\rho_i}{\rho_k} B_{ik}\right)^{-1}, \qquad (1)$$

where $B_{ik} = m_i(\mathbf{x})/m_k(\mathbf{x})$ is the Bayes factor of model M_i to M_k, and ρ_k is the prior probability of model k. (See, e.g., Chipman, George, and McCulloch 2001; Berger and Pericchi 2001, 2004; Kass and Raftery 1995; Ghosh and Samanta 2001; Barbieri and Berger 2004; Robert 2001; George and McCulloch 1996; Madigan and Raftery 1994 for discussion of Bayesian model selection techniques.) As an alternative to the selection of a single model, a common approach within the Bayesian framework is that of model averaging, which incorporates model uncertainty in addition to parameter uncertainty. Here interest would be in some predictive density,

$$\pi(y|\mathbf{x}) = \int_{\mathcal{K}} \int_{\Theta_k} \pi(y|\boldsymbol{\theta}_k)\tilde{\pi}_k(\boldsymbol{\theta}_k|\mathbf{x})\,d\boldsymbol{\theta}_k\,dk,$$

with integration over both model and parameter space. There is a fantastic literature on the application of Bayesian methods for model uncertainty. General review articles such as those by Clyde and George (2004), Chipman et al. (2001), Clyde (1999a), and Hoeting, Madigan, Raftery, and Volinsky (1999) contain a wealth of information and references. Similarly, Andrieu, Doucet, and Robert (2004) reviewed the recent computational and technological advances relating to Bayesian analyses, and Müller and Quintana (2004) and Heikkinen (2003) summarized the current state of nonparametric Bayesian inference.

A typical Bayesian analysis based on the foregoing will encounter two related problems. First, the density $m_k(\mathbf{x})$ in general will be unavailable because of analytic intractability. Second, the number of candidate models, $|\mathcal{K}|$, will often be very large, prohibiting a brute-force calculation of $M_k(\mathbf{x})$ via (1). One of the more flexible and popular techniques used to overcome these problems is Markov chain Monte Carlo (MCMC) methodology. For the approximation of $m_k(\mathbf{x})$ in particular, these include various forms of Metropolis–Hastings samplers. (See, e.g., Robert and Casella 2004; Cappé and Robert 2000; Gilks, Richardson, and Spiegelhalter 1996 for a discussion of standard MCMC methods and implementation issues.) However, techniques capable of simultaneously considering a large number of candidate models did not become available until the mid-1990s.

Almost exactly a decade ago, Green (1995) recast the terms and definitions involved in the Metropolis–Hastings algorithm in a more rigorous manner. In particular, the idea of using a time reversibility condition for the transition kernel of a Markov chain to ensure convergence to the desired stationary distribution was extended to more general state spaces. In integral form, the detailed balance condition for a general transition kernel P and its invariant distribution π can be written as

$$\int_{(\mathbf{x},\mathbf{x}') \in \mathcal{A}\times\mathcal{B}} \pi(d\mathbf{x})P(\mathbf{x}, d\mathbf{x}') = \int_{(\mathbf{x},\mathbf{x}') \in \mathcal{A}\times\mathcal{B}} \pi(d\mathbf{x}')P(\mathbf{x}', d\mathbf{x}) \qquad (2)$$

for all Borel sets $\mathcal{A} \times \mathcal{B} \subset \Theta$ for a general state space Θ (see, e.g., Green 2001; Tierney 1998). The novelty of this reexpression was that the generality of the state space under consideration now included formulations that could encompass multiple models. One often considered instance is $\Theta = \bigcup_{k\in\mathcal{K}} \Theta_k \times \{k\}$, that is, a countable union of subspaces of possibly varying dimensionality. Via standard Metropolis–Hastings updates, this development enabled the implementation of Markov chains simultaneously spanning both parameter and model space, Θ, with stationary distribution π that is absolutely continuous in Θ_k for each $k \in \mathcal{K}$ with respect to the n_k-dimensional Lebesgue measure. As a result, the estimation of posterior model probabilities and other marginal densities of

Scott A. Sisson is Lecturer in Statistics, Department of Statistics, School of Mathematics, University of New South Wales, Sydney 2052, Australia (E-mail: *Scott.Sisson@unsw.edu.au*). The author thanks Y. Fan and two referees for many useful comments on an earlier draft of this article.

Journal of the American Statistical Association
September 2005, Vol. 100, No. 471, Review Article
DOI 10.1198/016214505000000664

interest is in theory easily obtainable, irrespective of the order of $\mathcal{K}$. The general class of Markov chains that admit transitions between states of differing dimension have since been developed further to achieve a broad family of interrelated sampling frameworks. Given their model-spanning nature, these have recently been termed *transdimensional Markov chains*.

This article aims to achieve three objectives: (1) to provide an accessible evaluation of the current state of the art in terms of the practical implementation of transdimensional sampling technologies; (2) to provide an overview of the most recent and most important developments in multimodel sampling frameworks, in terms of ongoing theoretical and methodological development; and (3) to suggest some perspective of residual open problems and highlight those facets that would strongly benefit from further research.

Although transdimensional sampling algorithms are now well documented in the literature, in the first part of this article (Sec. 2) we present both a brief overview of the main frameworks that have been developed in the last 10 years and an assessment of their impact and implementation, including an evaluation of the freely available software for this purpose. We also consider non-Bayesian applications, the efficient estimation of Bayes factors, and the highly important, although frequently overlooked, issue of convergence assessment.

In the second part of this article (Sec. 3), we examine one facet of transdimensional sampling schemes that holds considerable potential for future research: the development of algorithms with increasing degrees of efficiency and automation. Achieving this—one of the fundamental goals of modern sampling frameworks—would permit routine implementation of transdimensional samplers by nonexpert practitioners, perhaps via stand-alone software packages such as the popular *WinBuGS* suite (Gilks, Thomas, and Spiegelhalter 1992; Spiegelhalter, Thomas, Best, and Gilks 1996b). Recent developments have made great strides in this direction, providing advances in the areas of between model transitions in terms of both efficiency and constructing generic mappings, the extension of perfect sampling schemes to the transdimensional case, and progress in default prior specifications over joint model and parameter spaces.

2. TRANSDIMENSIONAL MARKOV CHAINS AND THEIR IMPLEMENTATION

A number of frameworks have been proposed since the mid-1990s that supplement or extend the existing fixed-dimension Monte Carlo sampling schemes to encompass across model stochastic simulation. Each scheme may be related to others in a conceptually straightforward manner, facilitating natural settings for sampler comparison.

2.1 Sampling Frameworks

Introducing the $\Theta = \bigcup_{k\in\mathcal{K}} \Theta_k \times \{k\}$ formulation of model space, Grenander and Miller (1994) proposed a sampling strategy based on continuous time jump-diffusion dynamics. Such a Markov process essentially jumps between parameter spaces (and therefore models) at random times, and between the jumps follows a diffusion process according to a Langevin stochastic differential equation indexed by time, t, satisfying

$$d\boldsymbol{\theta}_k^t = d\mathbf{B}_k^t + \frac{1}{2}\nabla \log \pi(\boldsymbol{\theta}_k^t)\, dt, \tag{3}$$

where $d\mathbf{B}_k^t$ denotes an increment of Brownian motion and ∇ denotes the vector of partial derivatives. In practice, (3) is approximated by a discrete-time version with a Metropolis–Hastings step to preserve the stationary distribution π (Roberts and Tweedie 1996). This method has found some application in signal processing and other Bayesian analyses (e.g., Miller, Srivastava, and Grenander 1995; Phillips and Smith 1996), but has in general been superseded by the more accessible reversible-jump sampler (Green 1995). In fact, correcting for the time-discretization approximation via the Metropolis–Hastings acceptance probability template, the dump-diffusion sampler can be shown to result in an implementation of the reversible-jump algorithm (Besag 1994).

The widely implemented reversible-jump sampler was introduced by Green (1995) in a Bayesian model determination setting. Tierney (1998) and Green (2003a) both provided interesting expositions on this theme. One reason for the popularity of this algorithm in particular is conceptual: The framework is a natural generalization of the standard Markov chain theory, lending it certain appeal. In general, assuming that detailed balance (2) is satisfied, we may denote the acceptance probability of a proposed between-model move from $(\boldsymbol{\theta}_k, k)$ in model M_k to the state $(\boldsymbol{\theta}'_{k'}, k')$ in model $M_{k'}$ to be $\min\{1, A[(\boldsymbol{\theta}_k, k) \to (\boldsymbol{\theta}'_{k'}, k')]\}$. Here

$$A[(\boldsymbol{\theta}_k, k) \to (\boldsymbol{\theta}'_{k'}, k')] = \frac{L_{k'}(\mathbf{x}|\boldsymbol{\theta}'_{k'})p_{k'}(\boldsymbol{\theta}'_{k'})\rho_{k'}q(k' \to k)q_{k'}(\mathbf{u}'_{k'})}{L_k(\mathbf{x}|\boldsymbol{\theta}_k)p_k(\boldsymbol{\theta}_k)\rho_k q(k \to k')q_k(\mathbf{u}_k)} \left|\frac{\partial g_{k\to k'}(\boldsymbol{\theta}_k, \mathbf{u}_k)}{\partial(\boldsymbol{\theta}_k, \mathbf{u}_k)}\right|,$$

where $\boldsymbol{\theta}'_{k'} = g_{k\to k'}(\boldsymbol{\theta}_k, \mathbf{u}_k)$ for a random vector $\mathbf{u}_k \sim q_k(\mathbf{u}; \boldsymbol{\psi}_k)$ with parameter vector $\boldsymbol{\psi}_k$, and $q(k \to k')$ is the probability of proposing to move from model M_k to $M_{k'}$. Here $g_{k\to k'} : \Theta_k \times \Theta_k^q \to \Theta_{k'}$ denotes a mapping of the state $(\boldsymbol{\theta}_k, k)$ together with the vector $\mathbf{u}_k$ to the state $(\boldsymbol{\theta}'_{k'}, k')$. The mapping satisfies $g_{k'\to k}(g_{k\to k'}(\boldsymbol{\theta}_k, \mathbf{u}_k), \mathbf{u}'_{k'}) = \boldsymbol{\theta}_k$ and requires that $n_k + d_k = n_{k'} + d_{k'}$, where d_k is the dimension of Θ_k^q (known as "dimension matching").

It must be noted that the reversible-jump algorithm is not limited to the countable set of models $\mathcal{M}$, although it is frequently presented in this context. In fact, one may implement the sampler without knowing the size of the model space beforehand (which may contribute somewhat to the popularity of the algorithm), although at least some knowledge of the model space is recommended for the construction of an efficient chain. The most common setting involving an uncountable number of models is in Bayesian nonparametrics, where both the number of basis functions and the functions themselves are free to vary—via fractional polynomial regression (Royston and Altman 1994) or free-knot splines, for example. (See Denison, Holmes, Mallick, and Smith 2002; DiMatteo, Genovese, and Kass 2001; Denison, Mallick, and Smith 1998; Smith and Kohn 1996 for useful instances of Bayesian nonparametric curve fitting.)

Providing an alternative to samplers designed for implementation on unions of model spaces, a number of approaches have developed conventional Markov chain technology on a product supermodel space, $\Theta^* = \mathcal{K} \times \bigotimes_{k\in\mathcal{K}} \Theta_k$, that encompasses all model spaces jointly, thereby circumventing the necessity of between-model transitions. Defining a composite parameter

vector, $\boldsymbol{\theta}^*$, consisting of a concatenation of all parameters under all models, Carlin and Chib (1995) proposed a formulation in which the posterior distribution for the composite model space was given by

$$\pi(k, \boldsymbol{\theta}^*|\mathbf{x}) \propto L_k\left(\mathbf{x}|\boldsymbol{\theta}^*_{\mathcal{I}_k}\right) p_k\left(\boldsymbol{\theta}^*_{\mathcal{I}_k}\right) p_k\left(\boldsymbol{\theta}^*_{\mathcal{I}_{-k}}|\boldsymbol{\theta}^*_{\mathcal{I}_k}\right) \rho_k,$$

where $\mathcal{I}_k$ and $\mathcal{I}_{-k}$ are index sets identifying and excluding the parameters $\boldsymbol{\theta}_k$ from $\boldsymbol{\theta}^*$. Here $\mathcal{I}_k \cap \mathcal{I}_{k'} = \emptyset$ for all $k \neq k'$, so that the parameters for each model are distinct. In practice, this setting requires that the size of the model space, $|\mathcal{K}|$, be finite, thereby restricting its range of application relative to (say) the reversible-jump algorithm. Even when this setting is appropriate, a number of impracticalities are apparent for even a moderately sized model space. Although some of the computation in sampling the full parameter vector, $\boldsymbol{\theta}^*$, may be avoided (Godsill 2003; Dellaportas, Forster, and Ntzoufras 2002; Green and O'Hagan 1998), this approach requires the definition of $p_k(\boldsymbol{\theta}^*_{\mathcal{I}_{-k}}|\boldsymbol{\theta}^*_{\mathcal{I}_k})$, termed pseudopriors. Although their specification is essentially arbitrary in terms of obtaining the desired marginal distributions, sampler performance is sensitive to their specification, introducing practical problems in terms of efficiency and tractability (see Godsill 2001, 2003; Green 2003a for a discussion). However, it is believed that, in contrast to the lack of memory of previously visited states inherent in the reversible-jump sampler, in the product space formulations (which contain a perfect memory), the information contained within $\boldsymbol{\theta}^*_{\mathcal{I}_{-k}}$ may be useful in generating efficient between-model transitions when in model M_k. This idea was exploited by Brooks, Guidici, and Roberts (2003c), and we consider it in more detail in Section 3.1.

Comparing the methods of Green (1995) and Carlin and Chib (1995), Godsill (2001) proposed a further generalization of the foregoing that achieves enhanced flexibility by permitting individual model parameter vectors to overlap arbitrarily; that is, the restriction that $\mathcal{I}_k \cap \mathcal{I}_{k'} = \emptyset$ for all $k \neq k'$ is relaxed. This may seem intuitive for, say, nested models. The approach of Godsill (2001) may also be shown to encompass the reversible-jump sampler, thereby providing a general platform to facilitate comparisons between all previously introduced algorithms.

An alternative approach to the aforementioned formulations is based on spatial birth-and-death processes, originally investigated by Preston (1977) and Ripley (1977). Stephens (2000) proposed observing particular transdimensional statistical problems in the guise of continuous-time abstract marked point processes (see also Geyer and Møller 1994). Finite-mixture modeling is one such setting with obvious interpretations for the birth and death of mixture components. Recent work by Cappé, Robert, and Rydén (2003) has shown that the sampler of Stephens (2000) may be considered a particular continuous-time–limiting version of a sequence of reversible-jump algorithms.

A number of illustrative comparisons of the foregoing frameworks can be found in the literature. Andrieu, Djurić, and Doucet (2001) and Dellaportas et al. (2002) contrasted reversible jump with the pseudoprior approach of Carlin and Chib (1995), with the former analysis also providing a brief exposition on jump diffusion methodology. Godsill (2001, 2003) provided insight into the associations between composite union and product state-space formulations, and, as mentioned earlier, Cappé et al. (2003) examined the relationship between the reversible-jump algorithm, the sampler of Stephens (2000), and more general birth-and-death samplers.

Transdimensional sampling algorithms have had an undoubted influence in both the statistical and mainstream research literature. Perhaps not surprising, given the nature of certain technological advances over recent years, is the number of genetical applications relative to other subject areas. Overall, one in every five citations of Green (1995) can be broadly classified as genetics-based research. In general, the large majority of areas in which transdimensional Markov chains have strongly benefited to date have tended to be computationally or biologically related. Accordingly, a high number of developmental and application studies can be found in the signal processing literature and the related fields of computer vision and image analysis. Epidemiologic and medical studies also feature strongly, and MCMC and model selection methodological advances are other obvious high-inclusion groupings. With very few exceptions, the overwhelming use of the reversible-jump algorithm—and, by implication, transdimensional samplers in general—has been concerned with the generic problem of model selection.

2.2 Finessing Transdimensionality

Implementation of transdimensional Markov chains typically involves simultaneous exploration of both model and parameter space. However, depending on the aim or the complexity of a multimodel analysis, using transdimensional simulation techniques may be somewhat heavy-handed, as reductions to fixed-dimensional simulations may occasionally be attained. In some Bayesian model selection settings, transdimensionality can be avoided if one is prepared to make certain assumptions regarding prior choice, such as conjugacy or objective prior specifications (Berger and Pericchi 2001). Under these settings, explicit expressions for posterior model probabilities, $m_k(\mathbf{x})$, may be available (e.g., Casella and Moreno 2002). Similarly, it is not uncommon to find situations where approximations to $m_k(\mathbf{x})$ may be acceptable. In these cases the only "parameter" of interest is the model indicator k.

A second framework occurs when the full (normalized) conditional distributions, $\tilde{\pi}_k(\boldsymbol{\theta}_k|\mathbf{x})$, for each model M_k are known in closed form. This also is not uncommon in many conjugate models, such as linear regression, regression and classification trees, decomposable Gaussian and discrete graphical models, and even exponential family models, where simulation from block full conditional distributions is feasible. If the random vector $\mathbf{u}_k \sim q(\mathbf{u}; \boldsymbol{\psi}_k) = \tilde{\pi}_{k'}(\boldsymbol{\theta}'_{k'}|\mathbf{x})$ is a draw directly from this conditional distribution and the proposal state $\boldsymbol{\theta}'_{k'}$ determined through the mapping $\boldsymbol{\theta}'_{k'} = g_{k\to k'}(\boldsymbol{\theta}_k, \mathbf{u}_k) = \mathbf{u}_k$, then the reversible-jump acceptance probability (2.2) reduces to

$$A[(\boldsymbol{\theta}_k, k) \to (\boldsymbol{\theta}'_{k'}, k')] = \frac{\rho_{k'} q(k' \to k) m_{k'}(\mathbf{x})}{\rho_k q(k \to k') m_k(\mathbf{x})},$$

which is independent of both current and proposed parameter states. The algorithm thereby becomes a fixed-dimensional sampler over the space of models (see, e.g., Clyde 1999b). It is (or it shortly will be) possible to implement both of these simplified simulation frameworks in the popular *WinBuGS* suite;

see Section 2.3. When full model conditionals are not available, transdimensionality still may be avoided by adopting any of the product space formulations (Brooks et al. 2003c; Godsill 2001; Carlin and Chib 1995).

2.3 Implementation

Although most Bayesian analyses are likely to be novel in some aspect, thereby raising the likelihood that custom-written code is required for their implementation, at the same time it is easily appreciable that certain model types recur with sufficient frequency that new analyses may be based loosely on them. Depending on the nature of an application, it may even be the case that software is already freely available on the web, thereby permitting routine implementation of such algorithms via the code of their authors.

Table 1 lists some of the resources currently available from the transdimensional sampling community; most fixed-dimensional analyses may be performed using the *WinBuGS* suite. The most well-supported implementations are Gaussian mixture algorithms of varying forms (Cappé et al. 2003; Richardson and Green 1997; Sisson and Fan 2004a) and a number of methods that finesse the transdimensional nature of variable selection analyses by integrating out the within-model parameters, $\boldsymbol{\theta}_k$, before the analysis. The resulting simulations may then be performed in *WinBuGS* (e.g., Katsis and Ntzoufras 2005; Ntzoufras 2002). Other useful algorithms implement multiple changepoint analyses Green (1995), the automatic sampler of Green (2003a) (discussed further in Sec. 3.3), and an R (available at *http://www.r-project.org*) implementation of the transdimensional algorithm proposed by Petris and Tardella (2003).

Currently in development for *WinBuGS* is a new "*Jump*" component for the implementation of reversible-jump sampling for models in which the full conditional distribution for the within-model parameters is available in closed form (Lunn, Best, and Whittaker 2004). As discussed in Section 2.2, using these full conditionals as proposal distributions, the resulting simplification in the acceptance probability is then independent of the proposed parameters (which accordingly require generation only *after* the proposed move has been accepted). Such an approach thereby permits an efficient transdimensional implementation within the *WinBuGS* framework that retains the full parameter vectors, which is useful in terms of (say) predictive inference.

Even when generic or model-specific software is unavailable, many articles in the literature provide illuminating details on implementation of the various samplers. Table 2 provides a brief selection of just some of these in a range of modeling scenarios, and the interested reader is directed to the details therein. In practice, however, it is apparent that most routine data analyses are performed using the reversible-jump algorithm, given its more accessible nature. This algorithm is certainly the most well understood and well developed of all transdimensional sampling algorithms.

Despite the general dominance of the reversible-jump sampler, there are a number of situations in which adoption of one of the alternative algorithms may provide a simpler or more intuitive implementation. For example, when the total number of models is relatively small, so that the length of the composite parameter vector $\boldsymbol{\theta}^*$ is moderate, then adopting one of the product-space formulations may be useful (e.g., Ntzoufras 2002). Similarly, the birth-and-death approach has found some

Table 1. Freely Available Software for the Implementation of Transdimensional Samplers, With URL and Primary Citation

Green (2003a) Richardson and Green (1997) Green (1995)	*AutoRJ:* Automated reversible-jump MCMC *Nmix:* Bayesian analysis of univariate normal mixtures *Cpt:* Bayesian multiple changepoint analysis for point processes Available from *http://www.stats.bris.ac.uk/∼peter*
Cappé et al. (2003)	*CT/RJ-Mix:* Continuous-time and reversible-jump samplers for Gaussian mixtures Available from *http://www.tsi.enst.fr/∼cappe/ctrj_mix/*
Ntzoufras (2002) Katsis and Ntzoufras (2005)	Variable/model selection using BuGS Available from *http://www.jstatsoft.org/index.php?vol=7*; *http://www.ba.aegean.gr/ntzoufras/papers/paper13.htm*
Petris and Tardella (2003)	*HI:* Simulation from distributions supported by complex hyperplanes Available from *http://cran.r-project.org/src/contrib/Descriptions/HI.html*
Sisson and Fan (2004a)	*RjDiag:* Convergence diagnostic for transdimensional point processes (see *Nmix*) Available from *http://www.maths.unsw.edu.au/∼scott/*
Hoeting et al. (1999)	Links to various Bayesian model averaging software Available from *http://www.research.att.com/∼volinsky/bma.html*
Spiegelhalter et al. (1996b) Gilks et al. (1992)	*WinBuGS:* Software package for the Bayesian analysis of complex models using MCMC Available from *http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml* and the "Jump" component to become available from *http://homepages.tesco.net/∼creeping_death/*

Table 2. Articles With Detailed Implementations of the Respective Samplers

Jump diffusion	
Model comparison	Phillips and Smith (1996)
Image segmentation	Han, Tu, and Zhu (2004)
Target recognition	Miller et al. (1995)
Reversible-jump	
Changepoint models	Green (2001), Fan and Brooks (2000)
Signal processing	Larocque and Reilly (2002), Andrieu et al. (2001)
Mixture models	Richardson and Green (1997)
QTL mapping	Waagepetersen and Sorensen (2001), Stephens and Fisch (1998)
DNA segmentation	Boys and Henderson (2001)
Product space	
Switching models	Kim and Nelson (2001)
Variable/model selection	Spiegelhalter, Thomas, Best, and Gilks (1996a), Katsis and Ntzoufras (2005)
Birth and death	
Mixture models	Stephens (2000), Marin, Bernal, and Wiper (2003), Hurn, Justel, and Robert (2003)
Hidden Markov models	Shi, Murray-Smith, and Titterington (2002)
Switching models	Soegner (2000)

application in model settings that may be more naturally expressed in the point-process setting (Stephens 2000; Cappé et al. 2003), although these tend to be problem-specific (see Table 2 for references to specific illustrations). But jump-diffusion methods are more easily conceived in the discrete time setting. As a consequence, they have tended to be superseded in their application by the reversible-jump framework. Despite occasional implementation in applied settings, in general product-space approaches have found their greatest utility for sampler developmental purposes given the insights they provide into the relationships between the different sampling algorithms. We examine some of the resulting technological improvements in Section 3.1.

With the theory of model-spanning sampling frameworks now well established in the statistical literature, the main driving force of research in this field concerns the nature of difficulties encountered in their implementation. Issues of efficiency and mixing translate over from fixed-dimensional sampling schemes, as do problems in assessing convergence, with the additional obstacle of a substantial increase in the complexity of the problem. Between-model transitions play a key role in this setting. A related matter is the necessary specification of tuning parameters and the form of between-model mappings $g_{k\to k'}$—features that we examine in closer detail in Section 3.

2.4 Exploring Model Space

When the number of candidate models is large, model selection methods are generally concerned with the maximization (or minimization) of model-ranking functionals according to a nondeterministic optimization process (e.g., George and McCulloch 1993; Chipman et al. 2001). As a means to automate model selection, Brooks, Friel, and King (2003a) (see also Andrieu et al. 2000) proposed an extension to the standard simulated annealing framework by constructing a transdimensional Markov chain with stationary distribution proportional to the Boltzmann distribution,

$$\mathcal{B}_T(\boldsymbol{\theta}_k, k) \propto \exp\left\{-\frac{f(\boldsymbol{\theta}_k, k)}{T}\right\}, \tag{4}$$

where $T \geq 0$ and $f(\boldsymbol{\theta}_k, k)$, $(\boldsymbol{\theta}_k, k) \in \Theta$, is a model-ranking function to be minimized. A stochastic annealing framework may then be defined by periodically decreasing the value of T according to some schedule while using the Markov chain to explore functional space. As $T \to 0$, the distribution (4) converges to a point mass at $(\boldsymbol{\theta}^*_{k^*}, k^*) = \arg\max f(\boldsymbol{\theta}_k, k)$. Assuming adequate chain mixing, the algorithm will thereby identify the model determined by f. Applied to ecological capture-recapture analyses, this methodology has achieved success in facilitating the classical model selection procedure according to the Akaike information criterion (AIC) (Sisson and Fan 2004b; King and Brooks 2004) by setting $f(\boldsymbol{\theta}_k, k) = -2\log L_k(\mathbf{x}|\boldsymbol{\theta}_k) + 2n_k$. In Bayesian analyses one natural choice would be to consider $f(\boldsymbol{\theta}_k, k) = M_k(\boldsymbol{\theta}_k)$—the posterior model probability (e.g., Clyde 1999a). Stochastic optimization techniques such as the foregoing are not limited to the classical model selection frameworks. Under Bayesian decision-theoretic settings where loss functions defined on variable dimensional space take nonstandard form (e.g., Sisson and Hurn 2004), there is an obvious benefit in adopting flexible optimization methods for the derivation of Bayes rules.

Unfortunately, selection of the model with the highest-scoring model-ranking functional need not necessarily be the most useful in terms of a range of criteria. In the Bayesian normal linear model framework, Barbieri and Berger (2004) discussed optimality conditions when a single model must be chosen for predictive purposes. In particular, they were unable to identify general conditions under which the optimal predictive model coincided with that of the highest posterior probability model (although this does occur in some cases), but instead concluded that such optimality theorems in fact existed for the median probability model, defined to consist of those variables whose posterior probability of inclusion is at least $1/2$. In the classical analysis setting, Sisson and Fan (2004b) derived a sequence of profile models $\mathcal{M}^{\mathcal{P}} \subset \mathcal{M}$ such that $M_p^{\mathcal{P}} \in \mathcal{M}^{\mathcal{P}}$ is defined to consist of those variables whose "probability" of inclusion is at least p, for all $p \in [0, 1]$, and then perform a reduced model search on this set. Here their aim was to avoid the determination of overfitted models, resulting from use of the AIC, to retain (in this case) biological interpretability. These two examples illustrate subtle concerns with the process of model selection. The latter of these is indicative of the problems faced in the realistic specification of the model-ranking functional f when the number of candidate models is large, in

analogy with meaningful prior specification in the Bayesian setting. The former example underlines that the procedure is not necessarily straightforward, even when natural model-ranking measures are available.

2.5 Assessing Convergence

Perfect sampling schemes aside (see Sec. 3.2), under the assumption that an acceptably efficient method of constructing a transdimensional sampler is available, one obvious prerequisite to inference is that the Markov chain converges to its equilibrium state. This is a contentious issue even for fixed-dimensional cases. A priori convergence bounds are in general difficult or impossible to determine; a posteriori convergence diagnostics assess necessary rather than sufficient indicators of chain convergence (see, e.g., Mengersen, Robert, and Guihenneuc-Joyaux 1999; Cowles and Carlin 1996 for comparative reviews). The transdimensional setting generates additional concerns—in particular, how one might assess convergence not only within each of a potentially large number of models, but also across models with respect to posterior model probabilities.

One natural approach to this would be to implement independent subchain assessments, both within models and for the model indicator. However, in isolation this would erroneously associate convergence of the full chain with those of multiple subchains, thereby generating the potential for underestimation of convergence time for the full density. For example, with the focus purely on model selection, Brooks, Giudici, and Philippe (2003b) proposed various ideas based on the sample path of the model indicator, k, under the assumption that replicated chains that have converged will generate similar posterior model probability estimates. The focus on the model indicator permits the application of a number of nonparametric hypothesis tests; the chi-squared and Kolmogorov–Smirnov tests are discussed in detail. In this manner, distributional assumptions of the models (but not the statistics) are circumvented, at a price of associating marginal with full density convergence.

Other sampler performance issues arise as sparsely realized models of low posterior probability will be poorly represented (if at all), presenting an obstacle to even marginal assessment (Brooks 1997). One further problem is that of credibly assessing chain convergence of complex equilibrium distributions defined over high-dimensional state spaces by a univariate statistic, however well chosen.

Given the involved difficulties, it is hardly notable that to date there have been relatively few diagnostics designed specifically for transdimensional samplers. One strategy, proposed by Brooks and Giudici (2000), requires the determination of model reparameterization such that as many model parameters as possible, $\boldsymbol{\theta}^+$, retain their interpretation for all models under consideration. These parameters may then be monitored to provide an indication of chain performance. Specifically, Brooks and Giudici (2000) suggested a two-way ANOVA decomposition of the variance of a functional, $h(\boldsymbol{\theta}^+)$, over multiple-chain replications. A similar approach was advocated by Castelloe and Zimmerman (2002), who addressed the observed sensitivity of the Brooks and Giudici (2000) method to imprecise sample means from rarely visited models. They argued that a single visit to a rare model in one chain should not overwhelmingly dominate the diagnostic, and accordingly developed an unbalanced two-way ANOVA, with weights constructed in proportion to the frequency of model visits. Castelloe and Zimmerman (2002) also extended their methodology to the multivariate (MANOVA) setting on the observation that monitoring several functionals of marginal parameter subsets is more robust than monitoring a single statistic.

Although both groups of authors identified useful statistics to monitor, and Castelloe and Zimmerman (2002) offered some innovation in surmounting a problem of label-switching, one immediate problem with this approach in general is the difficulty identifying the requisite parameter set. A lesser issue regards the extent of approximation induced by violations of the ANOVA assumptions of independence and normality. Even ignoring the underlying implication of marginal assessment, the issue of parameter selection is magnified when considering that even common parameters may change meaning from one model to another (e.g., Berger and Pericchi 1996). This leads naturally to statistics, h, based on fitted and predicted values of observations as the obvious choice in many cases, reducing the problem to the fixed-dimensional setting (Green 2003a).

Sisson and Fan (2004a) suggested a method that circumvents the transdimensional nature of the problem when the underlying model can be formulated in the marked point-process framework of Stephens (2000). Specifically, the differences in intensity functions between chain replicates are determined by statistics based on the distributions of point-to-nearest-neighbor distances, thereby permitting the direct comparison of parameter vectors of varying dimension and, as a result, naturally incorporating a measure of across model convergence. Because of the nature of the construction of such functionals, Sisson and Fan (2004a) were able to monitor an arbitrarily large number of them. Although this approach may have some appeal, it is limited by the need to construct the model in the marked point-process setting.

Given the spectrum of difficulties involved in the performance monitoring of transdimensional samplers, and the obligingly small suite of tried-and-tested diagnostic tools available for this task, the ever-increasing numbers of articles containing analyses based on these sampling methods is perhaps some cause for concern. The most common approach adopted in the literature using the reversible-jump sampler (e.g., Jackson and Sharples 2004; Bottolo, Consonni, Dellaportas, and Lijoi 2003; Salmenkivi, Kere, and Mannila 2002; Suchard, Weiss, and Sinsheimer 2001; Kasuya and Takagawa 2001) rests on monitoring those parameters that "retain their interpretation as the sampler moves between topologies and [which] may be used effectively to monitor how well the MCMC sampler is performing" (Suchard et al. 2001), despite the aforementioned difficulties in selecting such parameters and the questionable "effectiveness" that marginal monitoring may provide. These parameters, generally a small subset of the full parameter set, are then monitored using popular fixed-dimensional performance measures (e.g., Smith 2004; Brooks and Gelman 1998; Gelman and Rubin 1992; Geweke 1992), although in many cases this analysis is limited to a single diagnostic. Of course, there is an obvious danger in monitoring only a single diagnostic to evaluate sampler performance—as much so as relying on marginal assessment. For instance, the finite mixture of normals (i.e., reversible-jump) algorithm of Richardson and Green

(1997) is perhaps the most studied in terms of its performance via a number of diagnostics (Sisson and Fan 2004a; Brooks et al. 2003b; Brooks and Giudici 2000; Richardson and Green 1997). However, despite the general perception that the sampler is fairly efficient, there is less unanimity on exactly "when" convergence may have been attained, depending on the diagnostic implemented.

Although it is undeniable that the benefits for the practitioner in implementing transdimensional sampling schemes are immense, it would seem arguable that the practical importance of ensuring chain convergence is often overlooked. More charitably, perhaps, it is more likely the case that current technology is insufficiently advanced to permit a more rigorous default assessment of sampler convergence, and until this shortcoming is resolved, conscienceous practitioners will be obliged to manage with the best that is currently available.

2.6 Estimating Bayes Factors

When the number of candidate models, $|\mathcal{K}|$, is large, using transdimensional sampling algorithms to evaluate Bayes factors between competing models raises issues of efficiency. In moving from $(\boldsymbol{\theta}_k, k)$ in model M_k to $(\boldsymbol{\theta}_{k'}, k')$ in model $M_{k'}$, Bartolucci and Scaccia (2003) demonstrated that the expected value of the reversible-jump acceptance probability under the distribution $f_k(\boldsymbol{\theta}_k, \mathbf{u}_k) = \pi_k(\boldsymbol{\theta}_k|\mathbf{x})q_k(\mathbf{u}_k)$ is bounded above by $B_{k'k}\rho_{k'}/\rho_k$. Accordingly, if model M_k accounted for a large portion of posterior mass, then the reversible-jump algorithm will tend to persist in model M_k and visit others models rarely. As a consequence, the resulting estimates of Bayes factors based on model-visit proportions will tend to be inefficient (e.g., Han and Carlin 2001); that is, the auxiliary random process adopted for transitions between models increases the variability of the estimator. In contrast, individually estimating the marginal model probabilities $m_k(\mathbf{x})$ and $m_{k'}(\mathbf{x})$ (Chib and Jeliazkov 2001; Chib 1995) or their ratio (Mira and Nicholls 2004; Meng and Schilling 2002; Chen and Shao 1997) via independent fixed-dimension simulations is more precise, although impracticalities emerge when the model space is large.

In an interesting recent development, Bartolucci and Mira (2003) (see also Bartolucci and Scaccia 2003) proposed an extension to the bridge estimator for estimating the ratio of normalizing constants of two distributions (Meng and Wong 1996). These authors augment the state spaces of the two distributions in the exact manner that is implicit in the specification of the auxiliary variables $\mathbf{u}_k$ and $\mathbf{u}'_{k'}$ in the reversible-jump algorithm, so that the distributions of interest, $f(\boldsymbol{\theta}_k, \mathbf{u}_k)$ and $f(\boldsymbol{\theta}_{k'}, \mathbf{u}'_{k'})$, are defined on the same, but not necessarily nested, space. Accordingly, using realizations from these two distributions, possibly directly from the reversible-jump sampler, the estimator of Bartolucci and Mira (2003) essentially integrates out the auxiliary random process, which thereby depends on a property of the acceptance probabilities and consequently provides more efficient estimates.

3. FUTURE PERSPECTIVES: TOWARD AUTOMATION

Since their inception, there has been a concerted drive to design sampling algorithms, both fixed-dimensional and transdimensional, that require the minimal initialization overheads but achieve the maximum in efficiency. This effort has manifested itself in a number of different ways, but to a greater extent current research is focused on embedding varying degrees of automation into the sampling process. Whether through default prior specification and one-size-fits-all generic algorithms, methods that adapt chain performance to enhance efficiency during implementation, or attempts to circumvent issues of convergence entirely, it is clear that the benefits in achieving fully automatic transdimensional algorithms are immense for the statistical community as a whole, as well as for those in other disciplines who wish to use these methodologies.

3.1 Efficient Chain Construction

Their popularity, conceptual simplicity, and obvious utility notwithstanding, transdimensional samplers have gained a reputation for inefficiency in implementation and poor performance in general. This is perhaps a somewhat unfair assessment; it would seem credible that at least some techniques of chain construction could be determined that would generate acceptable performance given that, for example, the reversible-jump algorithm encompasses all Metropolis–Hastings methods for between-model state transitions (Green 2003a). Failure to achieve acceptable performance could be considered merely a result of poorly constructed between-model transitions or inappropriate tuning of proposal distributions. Some methods targeted specifically at improving the acceptance rate of between-model transitions include the multistep candidate-generating scheme of Al-Awadhi, Hurn, and Jennison (2004), and the delayed rejection sampling scheme of Green and Mira (2001). Mengersen and Robert (2003) described a self-avoiding population Monte Carlo scheme aimed at increasing exploration of the state space.

It should perhaps even be anticipated that implementation of a transdimensional sampling scheme may result in enhanced mixing, even when applied in a fixed-dimensional setting. In this case, sampling from a single model distribution with a more sophisticated machinery might be considered loosely analogous with the extra power gained with augmented state–space sampling methods. For example, in the case of a strongly multimodal posterior, Richardson and Green (1997) reported markedly superior mixing in examining the output of a reversible-jump algorithm conditioned on there being exactly three mixture components, in contrast to the output generated by the respective fixed-dimension sampler. Similarly, George, Mengersen, and Davis (1999) achieved improved chain performance in an analysis concerning the ordering of genetic markers—a fixed-dimensional model—although in this case the "birth" and "death" moves were preformed simultaneously, so that the dimension of the model remained constant. In a short study, Green (2003a) presented a discussion on which inferential circumstances may determine whether the adoption of a transdimensional sampler may be beneficial (although see Han and Carlin 2001 for an argument to suggest that transdimensional sampling may have a detrimental effect on efficiency).

One complicating factor arises when the number of candidate models is considerable. Here the state space of the between-model structure may become difficult to visualize, causing problems in aligning regions of high probability and thereby in constructing efficient proposal templates—although

this factor may be of less concern under specific model settings, such as nested models. In addition, the task of manually tuning between-model transition variables via repeated pilot runs of the chain can become laborious and quickly prohibitive. There is, therefore, a strong argument for continued research into the development of assisted or automated proposal generation for both standard Metropolis–Hastings methods (e.g., Roberts, Gelman, and Gilks 1997) and transdimensional sampling schemes in particular.

Recently, Brooks, Guidici, and Roberts (2003c) (see also Ehlers and Brooks 2003) introduced a number of methods to achieve the automatic scaling of the proposal density. For a proposed move from $(\boldsymbol{\theta}_k, k)$ in M_k to model $M_{k'}$, one technique is based on identifying the random vector "centering point," $c(\boldsymbol{\theta}_k) = g_{k\to k'}(\boldsymbol{\theta}_k, \mathbf{u}_k(\boldsymbol{\theta}_k))$, such that for some particular choice of proposal vector, $\mathbf{u}_k(\boldsymbol{\theta}_k)$, the current and proposed states are identical in terms of likelihood contribution; that is, $L_k(\mathbf{x}|\boldsymbol{\theta}_k) = L_{k'}(\mathbf{x}|\mathbf{c}(\boldsymbol{\theta}_k))$. Given the constraint on $\mathbf{u}_k(\boldsymbol{\theta}_k)$, the relevant scaling parameters are then obtained to ensure that $A[(\boldsymbol{\theta}_k, k) \to (\mathbf{c}(\boldsymbol{\theta}_k), k')] = 1$. For example, in the case of autoregressive regression, if $M_{k'}$ is a higher-order model than M_k, then we may have $\mathbf{u}_k(\boldsymbol{\theta}_k) = \mathbf{0}$ as $L_k(\mathbf{x}|\boldsymbol{\theta}_k) = L_{k'}(\mathbf{x}|(\boldsymbol{\theta}_k, 0))$. (See also Ntzoufras et al. 2003 for a centering method in the context of linear models.)

A similarly motivated method is based on a series of nth-order conditions (for $n \geq 1$), which require that for the proposed move, the nth derivative of the acceptance probability equals the zero vector at $\mathbf{c}(\boldsymbol{\theta}_k)$,

$$\nabla^n A\big[(\boldsymbol{\theta}_k, k) \to (\mathbf{c}(\boldsymbol{\theta}_k), k')\big] = \mathbf{0}. \tag{5}$$

Similar to the foregoing, appropriate values for the proposal parameter vector, $\boldsymbol{\psi}_k$, are determined via (5), such that are likely to generate acceptance ratios close to one within a region. In this manner, proposal parameters are adapted to the current state of the chain at each stage rather than relying on a constant proposal parameter vector for all state transitions. It can be shown that for a simple two-model case, the aforementioned conditions are optimal in terms of the capacitance of the algorithm (Lawler and Sokal 1988).

Brooks et al. (2003c) also proposed a second class of models based on augmenting the state space with an auxiliary set of state-dependent variables, $\mathbf{v}_k$, so that the state space of $\pi(\boldsymbol{\theta}_k, \mathbf{v}_k|\mathbf{x}) = \pi_k(\boldsymbol{\theta}_k|\mathbf{x})\tau_k(\mathbf{v}_k)$ is of constant dimension for all models $M_k \in \mathcal{M}$. Although this fixed dimensionality is later relaxed, there is obvious analog with the product-space frameworks of Carlin and Chib (1995) and Godsill (2001). By considering updates of $\mathbf{v}_k$ via a slowly mixing Markov chain with a Gaussian stationary distribution a temporal memory is induced that persists in the $\mathbf{v}_k$ from state to state. In this way, the idea behind the auxiliary variables is to assist in between-model proposal transitions in that some memory of previous model states is retained. The authors demonstrate that this approach can significantly enhance mixing compared with an unassisted reversible-jump algorithm.

There is an obvious utility in the foregoing and other approaches for generic proposal design. However, one caveat with the foregoing schemes is that they assume prior specification of the between-model mapping $g_{k\to k'}$, and consequently that interest lies primarily in maximizing the between-model acceptance probability. When the candidate models have a strong degree of mutual consistency—so that there exist well-defined functionals of parameters with consistent meaning across models, and when prior specifications regarding these functionals are also consistent across models—the decomposition of the proposal model $M_{k'}$ into prior and likelihood terms, as proposed by Brooks et al. (2003c), is likely to generate natural local mappings between models Green (2003b). This can be observed in moment-matching methods, such as the well-known "split-and-merge" move types. When mutual consistency between models is lacking, however, or where the mappings, $g_{k\to k'}$, are suboptimal or even wholly undetermined, there is a clear limitation in the degree of chain efficiency that may be realized, and it is difficult to envisage how such schemes may assist.

In this sense, although maximizing within-model acceptance probabilities through local structural proposals will be useful in a broad range of modeling scenarios, more moderate acceptance rates, which aim to balance comprehensive within- and between-model transitions, may offer enhanced efficiencies in more general settings. (A useful, and familiar analogy can be found in the fixed-dimensional Metropolis–Hastings setting, whereby it is trivial to ensure a near-100% acceptance rate, but at the expense of poor exploration of the state space; see, e.g., Gelman, Roberts, and Gilks 1996.) From this perspective, a combination of both structurally local and more global between-model move types that do not rely on structural knowledge of the models to specify the between-model mapping, may provide the optimum specification. The case of automatically generating global between-model proposals is examined further in Section 3.3.

3.2 Perfect Sampling

Publication of the seminal paper by Propp and Wilson (1996) (see also Fill 1998) on the subject of stochastic sampling schemes that aim to draw realizations *exactly* from the stationary distribution of a Markov chain (see, e.g., Dimakos 2001; Casella, Lavine, and Robert 2000 for a review) generated considerable interest in the statistical community, based on the hope that such frameworks would become panacea to issues of convergence. (A valuable perfect sampling resource is *http://dbwilson.com/exact/* maintained by D. Wilson.) Despite considerable research in this area, perfect sampling schemes have proved difficult to implement in all but the simplest of modeling situations. This, coupled with the nonmodularity of the proposed algorithms, has resulted in a general reticence to embrace them as a mainstream technology. The increasing resemblance of exact simulation methods to standard algorithms involving Metropolis–Hastings steps is even leading some researchers to the opinion that little may be further gained by such methods in terms of ease and speed of implementation Robert (2003). Given the enhanced problems associated with the convergence of transdimensional Markov chains (Sec. 2.5), an exact multimodel algorithm would appear even more desirable than in the fixed-dimensional setting.

To date, the only methodological approach proposed in which an extension to a multimodal setting is discussed explicitly (Brooks, Fan, and Rosenthal 2002; Møller and Nicholls 1999) is based on an extension of the simulated tempering algorithm of Geyer and Thompson (1995). Intuitively, the fixed-model case proceeds by moving through the augmented state

space of heated models, each indexed by a temperature $\tau \in \mathcal{T} = \{1, 2, \ldots, \tau_{\max}\}$ (for example), and at each stage proposing a move to the distribution, indexed by $\tau = \tau^*$, from which it is possible to sample directly. The smallest possible probability of this occurring at any stage, assuming an identical state space Θ^T for all tempering models, is given by

$$\epsilon = \inf_{\boldsymbol{\theta} \in \Theta^T, \tau \in \mathcal{T}} q(\tau \to \tau^*) \min\big\{1, A[(\boldsymbol{\theta}, \tau) \to (\boldsymbol{\theta}, \tau^*)]\big\}.$$

Assuming that all possible chains commencing at time $t = -\infty$ coalesce into a single chain in model τ^* with probability ϵ, the focus would then naturally be on the first instance when this occurred in the reverse-time chain from $t = 0, -1, -2, \ldots, -T$. By construction, $T \sim \text{Geometric}(\epsilon)$. Consequently, commencing a forward-time Markov chain sampler starting in model τ^* at time $t = -T$ will have the effect of generating a draw exactly from the stationary distribution at time $t = 0$. Extension to the multimodel setting is immediate, requiring only a substitution of the temperature-augmented models by the set of candidate models $\mathcal{M}$ indexed by $k \in \mathcal{K}$, from which it is assumed that there is one model, M_{k^*}, from which it is possible to sample from directly. The algorithm then proceeds as before, following a transdimensional sampling scheme, where ϵ is now given by

$$\epsilon = \inf_{(\boldsymbol{\theta}_k, k) \in \Theta, \mathbf{u}_k \in \mathcal{Q}_k} q(k \to k^*) \times \min\big\{1, A\big[(\boldsymbol{\theta}_k, k) \to (g(\boldsymbol{\theta}_k, \mathbf{u}_k), k^*)\big]\big\}, \quad (6)$$

the smallest possible probability moving to model M_{k^*} over all possible models and states $(\boldsymbol{\theta}_k, k) \in \Theta$ and random vectors $\mathbf{u}_k \sim q(\mathbf{u}; \boldsymbol{\psi}_k)$ with $\mathbf{u}_k \in \mathcal{Q}_k$.

Although this algorithm is theoretically intuitive and appealing, a number of issues restrict its implementation in the general setting, most notably surrounding determination of ϵ in (6). Problems arise in the sense that the model M_{k^*} may not be known a priori, nor the model, M_k, which results in realizing the infimum. In addition, ϵ, or a more computationally accessible approximation, $\epsilon' \leq \epsilon$, may so small as to make simulation infeasible. Although the situation may be partially simplified in certain cases, such as nested models, it remains to be demonstrated that the foregoing approach (or any other) may be efficiently implemented in a transdimensional setting.

Given the nature of the problems of chain mixing and efficiency, one approach to good sampler construction would be to identify relevant issues for a given analysis and develop situation-specific solutions. This may be less than ideal for the MCMC novice wishing to use transdimensional sampling methods, who may not have adequate knowledge of the necessary implementational details or other specifications beyond that of the actual model to be analyzed—naturally, this is not (yet) recommended in practice! An alternative approach might be to seek more general solutions, by developing methods that aim to circumvent such implementation issues. We now consider some generic schemes that aim to adopt just such an automatically calibrated system.

3.3 Generic Samplers

The problem of efficiently constructing between-model mapping templates, $g_{k \to k'}$, as outlined earlier may be approached from an alternative perspective. Rather than adapting the properties of a fixed proposal mechanism to maximize the probability of accepting model-jumping transitions (Sec. 3.1), one possibility would be to remove the reliance on a user-specified method (such as birth/death or split/merge transitions) altogether. Although such an attractive ideal currently remains on the research horizon, a number of preliminary approaches to the development of generic automatic sampling frameworks have been recently proposed.

A reversible-jump analogy of the random-walk Metropolis sampler of Roberts (2003) was proposed by Green (2003a). For each of a small number of models $\mathcal{M}$, suppose that estimates of the first- and second-order moments of $\boldsymbol{\theta}_k$ are known; denote these by $\boldsymbol{\mu}_k$ and $\mathbf{B}_k\mathbf{B}_k^T$. In proposing a move from $(\boldsymbol{\theta}_k, k)$ to model $M_{k'}$, a new parameter vector is generated by setting

$$\boldsymbol{\theta}'_{k'} = \begin{cases} \boldsymbol{\mu}_{k'} + \mathbf{B}_{k'}[\mathbf{R}\mathbf{B}_k^{-1}(\boldsymbol{\theta}_k - \boldsymbol{\mu}_k)]_1^{n_{k'}} & \text{for } n_{k'} < n_k \\ \boldsymbol{\mu}_{k'} + \mathbf{B}_{k'}\mathbf{R}\mathbf{B}_k^{-1}(\boldsymbol{\theta}_k - \boldsymbol{\mu}_k) & \text{for } n_{k'} = n_k \\ \boldsymbol{\mu}_{k'} + \mathbf{B}_{k'}\mathbf{R}\begin{pmatrix} \mathbf{B}_k^{-1}(\boldsymbol{\theta}_k - \boldsymbol{\mu}_k) \\ \mathbf{u}_k \end{pmatrix} & \text{for } n_{k'} > n_k, \end{cases}$$

where $[\cdot]_1^m$ denotes the first m components of a vector, $\mathbf{u}_k \sim q_k(\mathbf{u}; \boldsymbol{\psi}_k)$ is an $(n_{k'} - n_k)$-dimensional vector of random numbers, and $\mathbf{R}$ is a orthogonal matrix of order $\max\{n_k, n_{k'}\}$. If the marginals $\pi_k(\boldsymbol{\theta}_k|\mathbf{x}) \sim \text{MVN}(\boldsymbol{\mu}_k, \mathbf{B}_k\mathbf{B}_k^T)$ and $\mathbf{u}_k$ is a standard normal vector, then choosing an appropriate model proposal density of $q(k \to k')/q(k' \to k) = \pi_{k'}(\mathbf{x})/\pi_k(\mathbf{x})$ would ensure that the acceptance probability $\min\{1, A[(\boldsymbol{\theta}_k, k) \to (\boldsymbol{\theta}_{k'}, k')]$, with

$$A[(\boldsymbol{\theta}_k, k) \to (\boldsymbol{\theta}_{k'}, k')] = \frac{\pi_{k'}(\boldsymbol{\theta}'_{k'}|\mathbf{x})}{\pi_k(\boldsymbol{\theta}_k|\mathbf{x})} \frac{q(k' \to k)}{q(k \to k')} \frac{|\mathbf{B}_{k'}|}{|\mathbf{B}_k|} \times \begin{cases} q_k(\mathbf{u}_k) & \text{for } n_{k'} < n_k \\ 1 & \text{for } n_{k'} = n_k \\ 1/q_k(\mathbf{u}_k) & \text{for } n_{k'} > n_k, \end{cases}$$

would equal unity. That is, the sampler would automatically achieve the detailed balance condition and is the motivation for the author. The implication is that high transition probabilities may be achieved when the marginal distributions π_k are unimodal with first- and second-order moments given by $\boldsymbol{\mu}_k$ and $\mathbf{B}_k\mathbf{B}_k^T$. Green (2003a) discussed a number of modifications to this general framework and illustrates it via variable selection and changepoint problems.

A related sampler was proposed by Godsill (2003), who, in adopting standard Gaussian approximation arguments, suggested the proposal-generating mechanism

$$\boldsymbol{\theta}'_{k'} = \boldsymbol{\mu}_{k'} + \mathbf{B}_{k'}\mathbf{v}_{k'},$$

where $\mathbf{v}_{k'} \sim q^*_{k'}$ is an $n_{k'}$-dimensional standard normal vector, which has similar detailed balance properties to the Green (2003a) sampler. Although a detailed comparative study has not yet been implemented, differences between the two samplers emerge when the target densities are non-Gaussian, in which there is a trade-off between both the target distribution ratios, and the difference in variability between the ratio $q^*_{k'}(\mathbf{v}_{k'})/q^*_k(\mathbf{v}_k)$ and $q_k(\mathbf{u}_k)$, which is generally of lower dimension (Godsill 2003).

The foregoing samplers have a number of obvious restrictions. Primarily they involve knowledge of the first- and second-order moments of the parameters under each model, something

that may be difficult to attain for anything but a small number of models. Green (2003a) obtained estimates of these via pilot chains on each model. The assumption of unimodality is also an important factor; the departure of the proposal distribution from the true conditional density strongly affects the acceptance rate of the algorithm. On a (multimodal) changepoint analysis, Green (2003a) demonstrated a relative efficiency of 29% of the automatic sampler compared with the results of a standard reversible-jump sampler (Green 1995), but with reduced implementation time. Recent work by Hastie (2004) has examined a finite mixture of normals for this sampler, yielding the expected improvements over nonnormal marginals, although at the obvious increase in initial computational expense. Such caveats notwithstanding, the avoidance of the "necessity" of specifying a between-model parameter mapping $g_{k\to k'}$ is an important step in the development of future sampling technologies.

3.4 Automatic and Objective Prior Specification

One of the obvious problems with the Bayesian approach to inference, in contrast with the sampler-based mechanics considered until now, is prior elicitation. This is a particular problem in analyses when the number of candidate models is large, because a specification that accurately represents given prior knowledge for all parameters and all models is typically infeasible. The use of noninformative or improper priors is therefore attractive for this task, and in general forms the basis of an analysis. The drawback to model selection with such priors is that the Bayes factor is determined arbitrarily, leading to a number of alternative default methods, such as default *proper* priors (e.g., Zellner and Siow 1980), fractional Bayes factors (O'Hagan 1995), and intrinsic Bayes factors (Berger and Pericchi 1996). (See, e.g., Berger and Pericchi 2001 for an extensive discussion.)

One recent development of the intrinsic Bayes factor approach that is particularly attractive from a computational viewpoint is the expected-posterior prior approach of Pérez and Berger (2002). Assuming standard noninformative priors, $p_k^N(\boldsymbol{\theta}_k)$ for each model, the expected-posterior prior for $\boldsymbol{\theta}_k$ under m^* is defined as

$$p_k^*(\boldsymbol{\theta}_k) = \int p_k^N(\boldsymbol{\theta}_k|\mathbf{y}^*)m^*(\mathbf{y}^*)\,d\mathbf{y}^*, \tag{7}$$

where $\mathbf{y}^*$ denotes a minimal vector of "imaginary" training data, $p_k^*(\boldsymbol{\theta}_k|\mathbf{y}^*)$ is the posterior distribution of the parameters $\boldsymbol{\theta}_k$ given the training data, and m^* is a predictive measure for $\mathbf{y}^*$.

Assuming that the (possibly improper) default priors p_k^N can be determining automatically, construction of (7) requires only the specification of m^*, which itself may be potentially determined given the modeling situation. For instance, in the case of nested models, one such choice might be

$$m^*(\mathbf{y}^*) = \int L_1(\mathbf{y}^*|\boldsymbol{\theta}_1)p_1(\boldsymbol{\theta}_1)\,d\boldsymbol{\theta}_1,$$

the predictive density of the training data under the base model, M_1. Most significantly for the expected-posterior prior approach, the generated prior distributions for different models are appropriately compatible, and problems of impropriety are also avoided (see Pérez and Berger 2002 for full details). Perhaps what makes the foregoing prior specification particularly attractive is that due to the probabilistic nature of the specification (7), their use is particularly suited to incorporation in hierarchical Markov chain sampling frameworks (e.g., Pérez and Berger 2001). The imaginary training sample $\mathbf{y}^*$ is simply considered one of the unknown parameters in the simulation.

3.5 Adaptation

The ability of a Markov chain to satisfactorily traverse model space is further complicated given that models with high posterior probability need not exhibit structural similarities and correspondingly may reside in relatively disparate portions of model space (see also Sec. 3.3). This creates obvious problems for the construction of between-model proposals $q(k \to k')$ and parameter mappings, $g_{k\to k'}$, between which it is natural to consider model-transitions based on local perturbations, both for conceptual ease and for ensuring a reasonable likelihood of accepting the proposed move. A similar situation occurs in fixed-dimensional settings in the presence of a strongly multimodal posterior.

Most research to address this problem has been conducted in the fixed dimension under the generic label of "adaptive" MCMC, which seeks to use the full sample path of the Markov chain to construct an efficient proposal density during chain implementation. (See, e.g., Frigessi 2003 for a concise statement on the current state of the art.) Briefly, the main issues center on the extent to which the stationary distribution adheres to the desired posterior given the manner of adaptation. Care must be taken to not adapt too quickly or inconsistently, or the wrong target distribution may be attained, a result that is all to easily achieved (see, e.g., Atchade and Rosenthal 2003). Subject to assumptions of uniformly ergodic transition kernels and bounded state spaces, the adaptive algorithm of Haario, Saksman, and Tamminen (2001), which depends on the full history of the chain, can be directly shown to yield unbiased Monte Carlo estimates of the expectations of bounded functionals (see Andrieu and Moulines 2002; Atchade and Rosenthal 2003 for results in more general settings). In comparison, if so-called "regeneration points" exist—such as an independent sample drawn from the "hot" distribution in a simulated tempering algorithm (Tierney 1996; Brooks et al. 2002), or an atom in the state space—then the adaptation may be implemented at these times. The dependence on the full chain history is consequently mitigated, and the Markovian structure is preserved (Sahu and Zhigljavsky 2003; Gåsemyr 2003; Gilks, Roberts, and Sahu 1998). Frigessi (2003) suggested that there may be scope for development in adopting dth-order Markov chains whose stationary distribution may be slightly biased but that are mathematically more flexible than chains based on the full sample-path history. This could conceivably be extended to variable-length Markov chains (Bühlmann and Wyner 1999).

Although focus is currently on the fixed-dimensional problem, it can be easily envisaged that adaptive methods will eventually graduate to the transdimensional setting, permitting the construction of between-model proposals that increasingly resemble full-model conditionals. In situations where the transdimensional nature of an analysis may be avoided (Sec. 2.2), this is already possible. For example, in a Gaussian variable selection setting using intrinsic priors—a fixed-dimensional problem because posterior model probabilities are known—Casella

and Moreno (2002) suggested a method of ensuring that the model proposal density approximates the posterior model probability $q(k \to k') \approx M_{k'}(\mathbf{x})$ as the chain length $N \to \infty$. In this manner the suggested transitions density $q(k \to k')$ seeks to avoid the problems associated with local modes of model space by proposing candidate models approximately in proportion to their posterior probabilities.

4. DISCUSSION

One of the fundamental goals of transdimensional sampling frameworks is to achieve high degrees of both efficiency and automation. In addition to providing a survey of the past decade of progress toward this goal, in this article we have presented a discussion on the some of problems associated with attaining this objective and illustrated some of the most recent attempts to engage it. Each of the areas highlighted has considerable potential for further development and innovation.

Given the degree of implementational difficulty associated with such methodologies, the aspect of automation of default model and chain specifications is particularly important with regard to the extent to which transdimensional Markov chains have been embraced within a broad array of application areas across a wide variety of disciplines. As a fundamental rule, it is important that statistical techniques be accessible to those who wish to adopt them while stopping short of black-box implementations. Public domain software packages, such as *WinBuGS* and its open-source companion *Open-BuGS*, are increasingly instrumental in propagating new methodologies through disparate disciplines. An all-in-one software package implementing automated and generic sampling schemes would surely prove a formidable resource for Bayesian analyses.

It might be imagined that under such a scenario, specification of the desired models would made via a graphical user interface or scripting language, with the option of highlighting a checkbox labelled "adopt objective priors." The software would then determine the most efficient within- and between-model transitions and implement the sampler, possibly even determining before sampling commences whether a perfect or standard sampling scheme will offer the greatest efficiency for a given computational time. Although such a software package is currently some way from realization, some recent innovative works discussed in this article have taken small, but confident strides in this direction. In doing so, they have generated heightened interest and enthusiasm in this goal.

As a closing note, although we have focused on reversible Markov chains defined by satisfying the detailed balance condition, there is a small literature that suggests that nonreversible chains may offer improvements in efficiency not available to their more accessible reversible counterparts. For example, Diaconis, Holmes, and Neal (2000) showed that nonreversibility can lead to improvements over the diffusive behavior of simple Markov chain sampling schemes. Neal (2004) (see also Mira and Geyer 2000) demonstrated that any reversible Markov chain on a finite and irreducible state space may be used to construct a nonreversible Markov chain on a related state space with asymptotic variance at least as small as that using the reversible chain, although typically this will be much smaller. Neal concluded that "this construction demonstrates that nonreversible chains have a fundamental advantage over reversible chains for MCMC estimation. Research into better MCMC methods may therefore best be focused on nonreversible chains."

[Received March 2004. Revised January 2005.]

REFERENCES

Al-Awadhi, F., Hurn, M. A., and Jennison, C. (2004), "Improving the Acceptance Rate of Reversible-Jump MCMC Proposals," *Statistics and Probability Letters*, 69, 189–198.

Andrieu, C., de Freitas, J., and Doucet, A. (2000), "Reversible-Jump MCMC Simulated Annealing for Neural Networks," in *Uncertainty in Articial Intelligence*, eds. C. Boutilier and M. Goldszmidt, San Francisco, CA: Morgan Kaufmann, pp. 11–18.

Andrieu, C., Djurić, P. M., and Doucet, M. (2001), "Model Selection by MCMC Computation," *Signal Processing*, 81, 19–37.

Andrieu, C., Doucet, A., and Robert, C. P. (2004), "Computational Advances for and From Bayesian Analysis," *Statistical Science*, 19, 118–127.

Andrieu, C., and Moulines, E. (2002), "On the Ergodicity Properties of Some Adaptive MCMC Algorithms," technical report, University of Bristol.

Atchade, Y. F., and Rosenthal, J. S. (2003), "On Adaptive Markov Chain Monte Carlo Algorithms," technical report, University of Montreal.

Barbieri, M. M., and Berger, J. O. (2004), "Optimal Predictive Model Selection," *The Annals of Statistics*, 32, 870–897.

Bartolucci, F., and Mira, A. (2003), "Efficient Estimate of Bayes Factors From Reversible-Jump Output," Technical Report 33, Università dell'Insubria, Dipartimento di Economia [in Italian].

Bartolucci, F., and Scaccia, L. (2003), "A New Approach for Estimating the Bayes Factor," technical report, University of Perugia.

Berger, J., and Pericchi, L. (1996), "The Intrinsic Bayes Factor for Model Selection and Prediction," *Journal of the American Statistical Association*, 91, 109–122.

——— (2001), "Objective Bayesian Methods for Model Selection: Introduction and Comparison," in *Model Selection*, ed. P. Lahiri, Hayward, CA: IMS, pp. 135–207.

——— (2004), "Training Samples in Objective Bayesian Model Selection," *The Annals of Statistics*, 32, 841–869.

Besag, J. (1994), Discussion of "Representations of Knowledge in Complex Systems," by U. Grenander and M. I. Miller, *Journal of the Royal Statistical Society*, Ser. B, 56, 591–592.

Bottolo, L., Consonni, G., Dellaportas, P., and Lijoi, A. (2003), "Bayesian Analysis of Extreme Values by Mixture Modelling," *Extremes*, 6, 25–47.

Boys, R. J., and Henderson, D. A. (2001), "A Comparison of Reversible-Jump MCMC Algorithms for DNA Sequence Segmentation Using Hidden Markov Models," *Computer Science and Statistics*, 33, 35–49.

Brooks, S. P. (1997), Discussion of "On Bayesian Analysis of Mixtures With an Unknown Number of Components," by S. Richardson and P. J. Green, *Journal of the Royal Statistical Society*, Ser. B, 59, 774–775.

Brooks, S. P., Fan, Y., and Rosenthal, J. S. (2002), "Perfect Forward Simulation via Simulated Tempering," technical report, University of Cambridge.

Brooks, S. P., Friel, N., and King, R. (2003a), "Classical Model Selection via Simulated Annealing," *Journal of the Royal Statistical Society*, Ser. B, 65, 503–520.

Brooks, S. P., and Gelman, A. (1998), "General Methods for Monitoring Convergence of Iterative Simulations," *Journal of Computational and Graphical Statistics*, 7, 434–455.

Brooks, S. P., and Giudici, P. (2000), "MCMC Convergence Assessment via Two-Way ANOVA," *Journal of Computational and Graphical Statistics*, 9, 266–285.

Brooks, S. P., Giudici, P., and Philippe, A. (2003b), "On Non-Parametric Convergence Assessment for MCMC Model Selection," *Journal of Computational and Graphical Statistics*, 12, 1–22.

Brooks, S. P., Guidici, P., and Roberts, G. O. (2003c), "Efficient Construction of Reversible-Jump Markov Chain Monte Carlo Proposal Distributions" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 65, 3–39.

Bühlmann, P., and Wyner, A. J. (1999), "Variable Length Markov Chains," *The Annals of Statistics*, 27, 480–513.

Cappé, O., and Robert, C. P. (2000), "Markov Chain Monte Carlo: 10 Years and Still Running!" *Journal of the American Statistical Association*, 95, 1282–1286.

Cappé, O., Robert, C. P., and Rydén, T. (2003), "Reversible-Jump MCMC Converging to Birth-and-Death MCMC and More General Continuous Time Samplers," *Journal of the Royal Statistical Society*, Ser. B, 65, 679–700.

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

Casella, G., Lavine, M., and Robert, C. (2000), "Explaining the Perfect Sampler," preprint, Université Paris 9—Dauphine.

Casella, G., and Moreno, E. (2002), "Objective Bayesian Variable Selection," submitted.
Castelloe, J. M., and Zimmerman, D. L. (2002), "Convergence Assessment for Reversible-Jump MCMC Samplers," Technical Report 313, University of Iowa, Dept. of Statistics and Actuarial Science.
Chen, M. H., and Shao, Q. M. (1997), "Estimating Ratios of Normalizing Constants for Densities With Different Dimensions," *Statistica Sinica*, 7, 607–630.
Chib, S. (1995), "Marginal Likelihood From the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–1321.
Chib, S., and Jeliazkov, I. (2001), "Marginal Likelihood From the Metropolis–Hastings Output," *Journal of the American Statistical Association*, 96, 270–281.
Chipman, H., George, E., and McCulloch, R. E. (2001), "The Practical Implementation of Bayesian Model Selection," in *Model Selection*, ed. P. Lahiri, Hayward, CA: IMS, pp. 67–134.
Clyde, M. A. (1999a), "Bayesian Model Averaging and Model Search Strategies," in *Bayesian Statistic 6*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Oxford University Press, pp. 157–185.
——— (1999b), "Discussion of "Bayesian Model Averaging: A Tutorial," by J. A. Hoeting, D. Madigan, A. E. Rafrety, and C. T. Volinsky, *Statistical Science*, 14, 401–404; corrected version available at *http://www.stat.washington.edu/www/research/online/hoeting1999.pdf*.
Clyde, M. A., and George, E. I. (2004), "Model Uncertainty," *Statistical Science*, 19, 81–94.
Cowles, M. K., and Carlin, B. P. (1996), "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review," *Journal of the American Statistical Association*, 91, 883–904.
Dellaportas, P., Forster, J. J., and Ntzoufras, I. (2002), "On Bayesian Model and Variable Selection Using MCMC," *Statistics and Computing*, 12, 27–36.
Denison, D. G. T., Holmes, C. C., Mallick, B. K., and Smith, A. F. M. (2002), *Bayesian Methods for Nonlinear Classification and Regression*, New York: Wiley.
Denison, D. G. T., Mallick, B. K., and Smith, A. F. M. (1998), "Automatic Bayesian Curve Fitting," *Journal of the Royal Statistical Society*, Ser. B, 60, 333–350.
Diaconis, P., Holmes, S., and Neal, R. M. (2000), "Analysis of a Non-Reversible Markov Chain Sampler," *The Annals of Applied Probability*, 10, 726–752.
Dimakos, X. K. (2001), "A Guide to Exact Simulation," *International Statistical Review*, 69, 27–48.
DiMatteo, I., Genovese, C. R., and Kass, R. E. (2001), "Bayesian Curve-Fitting With Free-Knot Splines," *Biometrika*, 88, 1055–1071.
Ehlers, R. S., and Brooks, S. P. (2003), "Constructing General Efficient Proposals for Reversible-Jump MCMC," technical report, Federal University of Paraná, Dept. of Statistics.
Fan, Y., and Brooks, S. P. (2000), "Bayesian Modelling of Prehistoric Corbelled Domes," *Journal of the Royal Statistical Society*, Ser. D, 49, 339–354.
Fill, J. A. (1998), "An Interruptible Algorithm for Exact Sampling via Markov Chains," *The Annals of Applied Probability*, 8, 131–162.
Frigessi, A. (2003), "On Some Current Research in MCMC," in *Highly Structured Stochastic Systems*, eds. P. J. Green, N. L. Hjort, and S. Richardson, Oxford, U.K.: Oxford University Press, pp. 172–175.
Gåsemyr, J. (2003), "On an Adaptive Metropolis–Hastings Algorithm With Independent Proposal," *Scandinavian Journal of Statistics*, 30, 159–173.
Gelman, A., Roberts, G. O., and Gilks, W. G. (1996), "Efficient Metropolis Jumping Rules," in *Bayesian Statistics 5*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Valencia, Spain: Valencia University Press, pp. 599–607.
Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulations Using Multiple Sequences," *Statistical Science*, 7, 457–511.
George, A. W., Mengersen, K. L., and Davis, G. P. (1999), "A Bayesian Approach to Ordering Gene Markers," *Biometrics*, 55, 419–429.
George, E. I., and McCulloch, R. E. (1993), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881–889.
——— (1996), *Markov Chain Monte Carlo in Practice*, London: Chapman & Hall, pp. 203–214.
Geweke, J. (1992), "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in *Bayesian Statistics 4*, eds. J. M. Bernardo, A. F. M. Smith, A. P. Dawid, and J. O. Berger, New York: Oxford University Press, pp. 169–193.
Geyer, C. J., and Møller, J. (1994), "Simulation Procedures and Likelihood Inference for Spatial Point Processes," *Scandinavian Journal of Statistics*, 21, 359–373.
Geyer, C. J., and Thompson, E. A. (1995), "Annealing Markov Chain Monte Carlo With Applications to Ancestral Inference," *Journal of the American Statistical Association*, 90, 909–920.
Ghosh, J. K., and Samanta, T. (2001), "Model Selection—An Overview," *Current Science*, 80, 1135–1144.
Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (eds.) (1996), *Markov Chain Monte Carlo in Practice*, Boca Raton, FL: Chapman & Hall/CRC.
Gilks, W. R., Roberts, G. O., and Sahu, S. K. (1998), "Adaptive Markov Chain Monte Carlo Through Regeneration," *Journal of the American Statistical Association*, 93, 1045–1054.
Gilks, W. R., Thomas, D., and Spiegelhalter, D. J. (1992), "Software for the Gibbs Sampler," *Computing Science and Statistics*, 24, 439–448.
Godsill, S. J. (2001), "On the Relationship Between Markov Chain Monte Carlo Methods for Model Uncertainty," *Journal of Computational and Graphical Statistics*, 10, 1–19.
——— (2003), Discussion of "Trans-Dimensional Markov Chain Monte Carlo," by P. J. Green, in *Highly Structured Stochastic Systems*, eds. P. J. Green, N. L. Hjort, and S. Richardson, Oxford, U.K.: Oxford University Press, pp. 199–203.
Green, P. J. (1995), "Reversible-Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732.
——— (2001), "Primer on Markov Chain Monte Carlo," in *Complex Stochastic Systems*, eds. O. E. Barndorff-Nielsen, D. R. Cox, and C. Klüppelberg, Boca Raton, FL: Chapman & Hall/CRC, pp. 1–62.
——— (2003a), "Trans-Dimensional Markov Chain Monte Carlo," in *Highly Structured Stochastic Systems*, eds. P. J. Green, N. L. Hjort, and S. Richardson, Oxford, U.K.: Oxford University Press, pp. 179–198.
——— (2003b), Discussion of "On Non-Parametric Convergence Assessment for MCMC Model Selection," by S. P. Brooks, P. Giudici, and G. O. Roberts, *Journal of the Royal Statistical Society*, Ser. B, 65, 48–49.
Green, P. J., and Mira, A. (2001), "Delayed Rejection in Reversible-Jump Metropolis–Hastings," *Biometrika*, 88, 1035–1053.
Green, P. J., and O'Hagan, A. (1998), "Model Choice With MCMC on Product Spaces Without Using Pseudo-Priors," Technical Report 98-01, Nottingham University, Dept. of Statistics.
Grenander, U., and Miller, M. I. (1994), "Representations of Knowledge in Complex Systems" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 56, 549–603.
Haario, H., Saksman, E., and Tamminen, J. (2001), "An Adaptive Metropolis Algorithm," *Bernoulli*, 7, 223–242.
Han, C., and Carlin, B. P. (2001), "MCMC Methods for Computing Bayes Factors: A Comparative Review," *Journal of the American Statistical Association*, 96, 1122–1132.
Han, F., Tu, Z., and Zhu, S.-C. (2004), "Range Image Segmentation by an Effective Jump-Diffusion Method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26, 1138–1153.
Hastie, D. (2004), "Developments in Markov Chain Monte Carlo," unpublished doctoral thesis, University of Bristol.
Heikkinen, J. (2003), "Trans-Dimensional Bayesian Non-Parametrics With Spatial Point Processes," in *Highly Structured Stochastic Systems*, eds. P. J. Green, N. Hjort, and S. Richardson, Oxford, U.K.: Oxford University Press, pp. 203–206.
Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999), "Bayesian Model Averaging: A Tutorial" (with discussion), *Statistical Science*, 14, 382–417; corrected version available at *http://www.stat.washington.edu/www/research/online/hoeting1999.pdf*.
Hurn, M., Justel, A., and Robert, C. P. (2003), "Estimating Mixtures of Regressions," *Journal of Computational and Graphical Statistics*, 12, 55–79.
Jackson, C. H., and Sharples, L. D. (2004), "Models for Longitudinal Data With Censored Changepoints," *Journal of the Royal Statistical Society*, Ser. C, 53, 149–162.
Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–796.
Kasuya, M., and Takagawa, I. (2001), "Model Uncertainty of Real Exchange Rate Forcast Over Mid-Term Horizons," Technical Report 01-23, Bank of Japan, Research and Statistics Department.
Katsis, A., and Ntzoufras, I. (2005), "Bayesian Hypothesis Testing for the Distribution of Insurance Claim Counts Using the Gibbs Sampler," *Journal of Computational Methods in Science and Engineering*, to appear.
Kim, C.-J., and Nelson, C. R. (2001), "A Bayesian Approach to Testing for Markov-Switching in Univariate and Dynamic Factor Models," *International Economic Review*, 42, 989–1013.
King, R., and Brooks, S. P. (2004), "A Classical Study of Catch-Effort Models for Hector's Dolphins," *Journal of the American Statistical Association*, 99, 325–333.
Larocque, J. R., and Reilly, J. P. (2002), "Reversible-Jump MCMC for Joint Detection and Estimation of Directions of Arrival in Coloured Noise," *IEEE Transactions on Signal Processing*, 50, 231–240.
Lawler, G., and Sokal, A. (1988), "Bounds on the L^2 Spectrum for Markov Chains and Markov Processes," *Transactions of the American Mathematical Society*, 309, 557–580.
Lunn, D. J., Best, N., and Whittaker, J. (2004), "Reversible Jump and Its Implementation in WinBuGS," technical report, Imperial College School of Medicine, London, Dept. of Epidemiology and Public Health.

Madigan, D., and Raftery, A. E. (1994), "Model Selection and Accounting for Uncertainty in Graphical Models Using Occam's Window," *Journal of the American Statistical Association*, 89, 1535–1546.

Marin, J. M., Bernal, M. R. R., and Wiper, M. P. (2003), "Using Weibull Mixture Distributions to Model Heterogeneous Survival Data," Technical Report 03-32, Carlos III University of Madrid, Dept. of Statistics and Economics.

Meng, X. L., and Schilling, S. (2002), "Wrap Bridge Sampling," *Journal of Computational and Graphical Statistics*, 11, 552–586.

Meng, X. L., and Wong, W. H. (1996), "Simulating Ratios of Normalising Constants via a Simple Identity: A Theoretical Exploration," *Statistica Sinica*, 6, 831–860.

Mengersen, K. L., and Robert, C. P. (2003), "IID Sampling Using Self-Avoiding Population Monte Carlo: The Pinball Sampler," in *Bayesian Statistics 7*, eds. J. M. Bernardo, M. J. Baryarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West, Oxford, U.K.: Oxford University Press, pp. 277–292.

Mengersen, K. L., Robert, C. P., and Guihenneuc-Joyaux, C. (1999), "MCMC Convergence Diagnostics: A Review," in *Bayesian Statistics 6*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Oxford University Press, pp. 415–440.

Miller, M. I., Srivastava, A., and Grenander, U. (1995), "Conditional-Mean Estimation via Jump-Diffusion Processes in Multiple Target Tracking/Recognition," *IEEE Transactions on Signal Processing*, 43, 2678–2690.

Mira, A., and Geyer, C. J. (2000), "On Non-Reversible Markov Chains," in *Fields Institute Communications: Monte Carlo Methods*, Vol. 26, ed. N. Madras, American Mathematical Society, pp. 93–108.

Mira, A., and Nicholls, G. (2004), "Bridge Estimation of the Probability Density at a Point," *Statistica Sinica*, 14, 603–612.

Møller, J., and Nicholls, G. K. (1999), "Perfect Simulation for Sample-Based Inference," technical report, Aalborg University.

Müller, P., and Quintana, F. A. (2004), "Nonparametric Bayesian Data Analysis," *Statistical Science*, 19, 95–110.

Neal, R. M. (2004), "Improving Asymptotic Variance of MCMC Estimators: Non-Reversible Chains Are Better," Technical Report 0406, University of Toronto, Dept. of Statistics.

Ntzoufras, I. (2002), "Gibbs Variable Selection Using BuGS," *Journal of Statistical Software*, 7.

Ntzoufras, I., Dellaportas, P., and Forster, J. J. (2003), "Bayesian Variable and Link Determination for Generalised Linear Models," *Journal of Statistical Planning and Inference*, 111, 165–180.

O'Hagan, A. (1995), "Fractional Bayes Factors for Model Comparisons," *Journal of the Royal Statistical Society*, Ser. B, 57, 99–138.

Pérez, J. M., and Berger, J. O. (2001), "Analysis of Mixture Models Using Expected Posterior Priors, With Application to Classification of Gamma Ray Bursts," in *Bayesian Methods With Applications to Science, Policy and Official Statistics*, eds. E. George and P. Nanopolus, Luxembourg: Official Publications of the European Communities, pp. 401–410.

——— (2002), "Expected Posterior Prior Distributions for Model Selection," *Biometrika*, 89, 491–511.

Petris, G., and Tardella, L. (2003), "A Geometric Approach to Transdimensional Markov Chain Monte Carlo," *The Canadian Journal of Statistics*, 31, 469–482.

Phillips, D. B., and Smith, A. F. M. (1996), "Bayesian Model Comparison via Jump Diffusions," in *Markov Chain Monte Carlo in Practice*, eds. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, London: Chapman & Hall, pp. 215–239.

Preston, C. J. (1977), "Spatial Birth-and-Death Processes," *Bulletin of the International Statistical Institute*, 46, 371–391.

Propp, J. G., and Wilson, D. B. (1996), "Exact Sampling With Coupled Markov Chains and Applications to Statistical Mechanics," *Random Structures and Algorithms*, 9, 223–252.

Richardson, S., and Green, P. J. (1997), "On Bayesian Analysis of Mixtures With an Unknown Number of Components" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 59, 731–792.

Ripley, B. D. (1977), "Modelling Spatial Patterns" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 39, 172–212.

Robert, C. (2001), *The Bayesian Choice* (2nd ed.), New York: Springer-Verlag.

——— (2003), "Advances in MCMC: A Discussion," in *Highly Structured Stochastic Systems*, eds. P. J. Green, N. Hjort, and S. Richardson, Oxford, U.K.: Oxford University Press, pp. 167–171.

Robert, C., and Casella, G. (2004), *Monte Carlo Statistical Methods* (2nd ed.), New York: Springer-Verlag.

Roberts, G. O. (2003), "Linking Theory and Practice of MCMC," in *Highly Structured Stochastic Systems*, eds. P. J. Green, N. Hjort, and S. Richardson, Oxford, U.K.: Oxford University Press, pp. 145–166.

Roberts, G. O., Gelman, A., and Gilks, W. R. (1997), "Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms," *The Annals of Applied Probability*, 7, 110–120.

Roberts, G. O., and Tweedie, R. L. (1996), "Exponential Convergence of Langevin Diffusions and Their Discrete Approximations," *Bernoulli*, 2, 341–364.

Royston, P., and Altman, D. G. (1994), "Regression Using Fractional Polynomials of Continuous Covariates: Parsimonious Parametric Modelling," *Applied Statistics*, 43, 429–453.

Sahu, S. K., and Zhigljavsky, A. A. (2003), "Self-Regenerative Markov Chain Monte Carlo With Adaption," *Bernoulli*, 9, 395–422.

Salmenkivi, M., Kere, J., and Mannila, H. (2002), "Genome Segmentation Using Piecewise Constant Intensity Models and Reversible-Jump MCMC," *Bioinformatics*, 18, S211–S218.

Shi, J. Q., Murray-Smith, R., and Titterington, D. M. (2002), "Birth–Death MCMC Methods for Mixtures With an Unknown Number of Components," Technical Report TR-2002-117, University of Glasgow, Dept. of Computing Science.

Sisson, S. A., and Fan, Y. (2004a), "A Distance-Based Diagnostic for Trans-Dimensional Markov Chains," technical report, University of New South Wales, School of Mathematics.

——— (2004b), "Towards Automating Model Selection for a Mark-Recapture-Recovery Analysis," technical report, University of New South Wales, School of Mathematics.

Sisson, S. A., and Hurn, M. A. (2004), "Bayesian Point Estimation of Quantitative Trait Loci," *Biometrics*, 60, 60–68.

Smith, B. J. (2004), "Bayesian Output Analysis Program (BOA), User's Manual, Version 1.1," technical report, University of Iowa, Dept. of Biostatistics, available at *http://www.public-health.uiowa.edu/boal.html.*

Smith, M., and Kohn, R. (1996), "Nonparametric Regression Using Bayesian Variable Selection," *Journal of Econometrics*, 75, 317–344.

Soegner, L. (2000), "Selection of the Number of States by Birth–Death Processes," technical report, Vienna University of Economics and Business Administration, Dept. of Economics.

Spiegelhalter, D., Thomas, A., Best, N., and Gilks, W. (1996a), *BUGS 0.5: Examples*, Vol. 2, Cambridge, U.K.: MRC Biostatistics Unit, Institute of Public Health, available at *http://www.mrc-bsu.cam.ac.uk/bugs/documentation/exampvol2/node20.html*.

——— (1996b), *BUGS: Bayesian Inference Using Gibbs Sampling*, Version 0.50, Cambridge, U.K.: Medical Research Council Biostatistical Unit.

Stephens, D. A., and Fisch, R. D. (1998), "Bayesian Analysis of Quantitative Trait Locus Data Using Reversible-Jump Markov Chain Monte Carlo," *Biometrics*, 54, 1334–1347.

Stephens, M. (2000), "Bayesian Analysis of Mixture Models With an Unknown Number of Components: An Alternative to Reversible-Jump Methods," *The Annals of Statistics*, 28, 40–74.

Suchard, M. A., Weiss, R. E., and Sinsheimer, J. S. (2001), "Bayesian Selection of Continuous-Time Markov Chain Evolutionary Models," *Molecular Biology and Evolution*, 18, 1001–1013.

Tierney, L. (1996), "Introduction to General State–Space Markov Chain Theory," in *Markov Chain Monte Carlo in Practice*, eds. W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, London: Chapman & Hall, pp. 59–74.

——— (1998), "A Note on Metropolis–Hastings Kernels for General State Spaces," *The Annals of Applied Probability*, 8, 1–9.

Waagepetersen, R., and Sorensen, D. (2001), "A Tutorial on Reversible-Jump MCMC With a View Toward Applications on QTL Mapping," *International Statistical Review*, 69, 49–62.

Zellner, A., and Siow, L. (1980), "Posterior Odds for Selected Regression Hypotheses," in *Bayesian Statistics*, eds. D. V. L. J. M. Bernardo, M. H. DeGroot, and A. F. M. Smith, Valencia: Valencia University Press, pp. 585–603.