Of Bugs and Birds: Markov Chain Monte Carlo for Hierarchical Modeling in Wildlife Research
Author(s): William A. Link, Emmanuelle Cam, James D. Nichols, Evan G. Cooch
Source: *The Journal of Wildlife Management*, Vol. 66, No. 2 (Apr., 2002), pp. 277-291
Published by: Allen Press
Stable URL: http://www.jstor.org/stable/3803160
Accessed: 17/01/2010 17:49

OF *BUGS* AND BIRDS: MARKOV CHAIN MONTE CARLO FOR HIERARCHICAL MODELING IN WILDLIFE RESEARCH

WILLIAM A. LINK,[1] USGS Patuxent Wildlife Research Center, 11510 American Holly Drive, Laurel, MD 20708, USA
EMMANUELLE CAM,[2] USGS Patuxent Wildlife Research Center, 11510 American Holly Drive, Laurel, MD 20708, USA
JAMES D. NICHOLS, USGS Patuxent Wildlife Research Center, 11510 American Holly Drive, Laurel, MD 20708, USA
EVAN G. COOCH, Department of Natural Resources, Fernow Hall, Cornell University, Ithaca, NY 14853, USA

Abstract: Markov chain Monte Carlo (MCMC) is a statistical innovation that allows researchers to fit far more complex models to data than is feasible using conventional methods. Despite its widespread use in a variety of scientific fields, MCMC appears to be underutilized in wildlife applications. This may be due to a misconception that MCMC requires the adoption of a subjective Bayesian analysis, or perhaps simply to its lack of familiarity among wildlife researchers. We introduce the basic ideas of MCMC and software *BUGS* (Bayesian inference using Gibbs sampling), stressing that a simple and satisfactory intuition for MCMC does not require extraordinary mathematical sophistication. We illustrate the use of MCMC with an analysis of the association between latent factors governing individual heterogeneity in breeding and survival rates of kittiwakes (*Rissa tridactyla*). We conclude with a discussion of the importance of individual heterogeneity for understanding population dynamics and designing management plans.

JOURNAL OF WILDLIFE MANAGEMENT 66(2):277–291

Key words: Bayesian analysis, kittiwake, Markov chain Monte Carlo, *Rissa tridactyla*, statistical methods.

This paper is an introduction to Markov chain Monte Carlo (MCMC), a powerful statistical tool that is used to analyze large, complicated data sets, especially those with complex hierarchical structures. The basic ideas of MCMC were introduced almost 50 years ago (Metropolis et al. 1953) and gained popularity during the 1980s in image processing (Geman and Geman 1984). A growing appreciation of the usefulness of MCMC has led to an explosion of publications in the statistical literature (Gilks et al. 1996). However, relatively few examples are found in wildlife-related applications.

We suggest several reasons why MCMC has not been more widely used in wildlife applications. First, MCMC has a decidedly Bayesian flavor, which may not appeal to data analysts with a classical (Frequentist) training. However, MCMC can be used as a tool to obtain the maximum likelihood estimates used by Frequentists, even for models with complexity that defies conventional analysis. Markov chain Monte Carlo also can be used in Objective Bayes analyses, the results of which are similar to those of Frequentist analyses. We begin with a brief review of Bayesian modeling, contrasting it with the Frequentist approach. This review lays the foundation for description of MCMC and describes Objective Bayes methods.

Another reason MCMC is not yet widely used among wildlife biologists may be a lack of familiarity. Markov chain Monte Carlo involves some complex mathematical ideas. A need exists for a clear accounting of what can be done with it and how it works. We describe the need for MCMC and its basic ideas and mechanisms. A simple and satisfactory intuition for MCMC does not require extraordinary mathematical sophistication.

We illustrate the usefulness of MCMC by analyzing the association between latent factors governing individual heterogeneity in breeding and survival rates of kittiwakes. The data set consists of survival and breeding records for 845 birds, collected over 13 years. The question of interest was whether there are trade-offs between components of fitness. For example, is it true that better breeders tend to have lower survival rates? The models we fit to these data are quite complex—impossible to fit using conventional methods—but are satisfactorily estimated using MCMC. We illustrate the application of MCMC to these data, using software *BUGS* (Spiegelhalter et al. 1995), available for free download (http://www.mrc-bsu.cam.ac.uk/bugs/).

FREQUENTIST, BAYES, AND OBJECTIVE BAYES MODELS

The primary distinction between Bayesian and Frequentist analyses is in the interpretation of model parameters. In both types of analysis, data (Y) are regarded as sampled from a sampling distrib-

[1] E-mail: william_link@usgs.gov
[2] Present address: Laboratoire de Biologie, Université de Bretagne Occidentale, 29285 Brest Cedex, France.

ution $f(Y|\theta)$, governed by parameter θ. In a Frequentist analysis, θ is regarded as a fixed, unknown quantity, while in a Bayesian analysis, θ is regarded as a random variable. In this section, we describe some of the consequences of this distinction.

Frequentist analyses describe the effects of variability on data by reference to hypothetical replicate data sets. These hypothetical replicates are drawn from the same sampling distribution, or a hypothesized version of the sampling distribution, that produced the data. Familiar statistical concepts such as *P*-values and unbiasedness are defined with reference to such hypothetical replicate draws. Statistical procedures are evaluated by their typical performance under similar circumstances.

An important difference exists between an estimator and an estimate. An estimator can be thought of as a machine that produces estimates; accuracy and precision are properties of the machine rather than of individual products. Thus, it is incorrect to describe an estimate as unbiased: unbiasedness is a property of estimators, summarized over applications to hypothetical replicate data sets. Confidence interval coverage rates are similarly defined, as features of the machine rather than of the product. The correct interpretation of confidence interval coverage rates is conveyed by analogy to a 95% elephant gun, fired over the shoulder while running away: you don't know whether it has stopped the charging beast, but there is some comfort in knowing that the desired result is obtained in 95% of similar circumstances.

In a Bayesian analysis, the parameter θ is regarded as a random variable. This paradigm has the consequence that direct probability statements are made about the parameters themselves, in contrast to the more indirect statements of Frequentist analyses.

Two probability distribution functions are used to make statements about θ, called the prior and posterior distributions. That the same random variable can be described by 2 different distributions may seem strange, but the concept is not at all unfamiliar: one might say that the chance the home team will win today is 50%; however, if the manager chooses to start the ace pitcher, the chance increases to 70%. The additional information modifies the probabilities. This is the case with descriptions of parameters in a Bayesian model. The prior distribution (or simply, the prior) summarizes what is known about the ranges and associated probabilities for θ without reference to the data Y. The posterior distribution (the posterior) provides the same summaries, but as informed by the data Y.

In a Bayesian analysis, inference about θ is based on the posterior distribution $f(\theta|Y)$, obtained by applying Bayes' theorem to the prior $\pi(\theta)$, and the sampling distribution $f(Y|\theta)$, by means of the calculation

$$f(\theta|Y) = \frac{f(Y|\theta)\pi(\theta)}{\int f(Y|\theta)\pi(\theta)\,d\theta}. \quad (1)$$

Thus, the posterior distribution describes the ranges of possible values for θ and their probabilities as indicated by the combination of data and prior knowledge. The Bayesian paradigm provides a formal mechanism for combining existing knowledge with indications provided by the data at hand, expressed informally as

$$Prior + Data = Posterior. \quad (2)$$

To give a concrete example, suppose that we wish to estimate the mean θ of a normal distribution based on a sample of n observations; for simplicity, assume that the variance σ^2 is known. A Frequentist analysis neither requires nor allows for prior knowledge about the likely range of values for θ, simply using $\bar{x}$, the sample mean, as its estimate for θ. A typical Bayesian analysis treats θ as a random variable, itself sampled from a normal prior with mean η and variance τ^2, both of which are assumed to be known. Under this model (the normal–normal mixture), the posterior distribution of θ also is a normal distribution, but with mean

$$E(\theta|data) = w\eta + (1 - w)\bar{x}, \quad (3)$$

and variance

$$Var(\theta|data) = (1 - w)\left(\frac{\sigma^2}{n}\right), \quad (4)$$

where

$$w = \frac{(\sigma^2/n)}{\tau^2 + (\sigma^2/n)}. \quad (5)$$

The posterior mean sometimes is referred to as a Bayes estimate, and written as $\hat{\theta}_B$.

Two important features of this example are characteristic of Bayesian estimation. The first is that $\hat{\theta}_B$ is a weighted average of the prior mean and the sample mean (3), with weights determined by the precision of existing knowledge (τ^2) relative to the new information provided by the data (i.e., $Var(\bar{x}) = (\sigma^2/n)$. Thus, Bayesian

estimation often is described as shrinkage toward a prior value: $\hat{\theta}_B$ is obtained by beginning at $\bar{x}$ and moving a portion of the way along a line toward the prior mean η. If $\bar{x}$ is a relatively imprecise estimate of θ, w will be close to 1, and the Bayesian estimate will be close to the prior mean. On the other hand, the greater the precision of $\bar{x}$, the closer w is to 0, and the closer the agreement between $\bar{x}$ and $\hat{\theta}_B$.

The second feature of interest in this example has to do with the mean squared error (MSE) of $\hat{\theta}_B$ as an estimator of θ. It can be shown that $\text{MSE}(\hat{\theta}_B) = Var(\theta|data)$, and since $0 \leq w \leq 1$, it follows from (4) that

$$\text{MSE}(\hat{\theta}_B) < \frac{\sigma^2}{n} = Var(\bar{x}) = \text{MSE}(\bar{x}). \qquad (6)$$

This reduction of mean squared error is a reflection of the increased knowledge brought to bear on the estimation by the prior knowledge of θ. This is a common feature of Bayesian analysis. The improvement is the consequence of a richer model through specification of the prior.

The Bayesian paradigm is useful for analysis of complex data sets that are governed by numerous parameters. For example, mark–recapture studies may involve hundreds of parameters, some governing survival rates, and others related to resighting rates. In such cases, it is desirable to examine patterns among parameters, such as the temporal trend in survival rates. This is sometimes done by supposing that the survival rates, suitably transformed, fall precisely on a given line. A more natural assumption is that the transformed survival rates more or less conform to a linear relation and are subject to additional temporal variation. The Bayesian approach, in which parameters are regarded as random variables, deals effectively with the existence of such pattern in parameters: the pattern of change is reflected in the means of the prior distributions (Link 1999).

In the foregoing example, we supposed that the prior distribution was known. It is often possible to relax this specification, to suppose only that the prior distribution is of a known family, governed by an unknown hyperparameter. We may then assign hyperprior distributions to the hyperparameters, and so on, so that a hierarchy of relations among parameters is established. At some point, however, Bayesian analysis must begin with a known distribution, and it is this specification of a prior that leads many classically trained analysts to reject the Bayesian paradigm as too subjective. The increased precision noted in (6) comes at a cost: the performance of the estimator $\hat{\theta}_B$ depends on the validity of the model specification, to which the Bayesian has added the specification of a prior distribution for θ.

What, it may be asked, prevents an unscrupulous data analyst from selecting a prior on the basis of the posterior it yields? There are 2 answers: first, that the prior distributions used should be included in the presentation of a Bayesian analysis. Indeed, it is instructive to try several different priors and to evaluate the relation between prior and posterior. We illustrate this principle in our subsequent analysis of kittiwake data. If the posterior distribution is highly sensitive to alternative (reasonable) choices of the prior, it is likely that the information content of the data is small.

Another response to the concern of subjectivity is to carry out an Objective Bayes analysis. Objective Bayes methods use prior distributions described as flat, vague, or noninformative; these terms sometimes are used in slightly different senses, but the basic idea is that the priors have been chosen so as to reflect only a very limited or imprecise prior knowledge of θ. An objective Bayes analysis amounts to substituting 0 for Prior in (2).

To illustrate, consider the foregoing example of estimating a normal mean: if the prior variance, τ^2, is large, the shrinkage factor w will be close to zero and the Bayes estimate, $\hat{\theta}_B$, will be close to the classical estimate $\bar{x}$. The posterior distribution $f(\theta|data)$ approximates a normal distribution with mean of $\bar{x}$ and variance σ^2/n, so that the usual Frequentist confidence interval

$$\bar{x} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

is legitimately interpreted as having probability $(1 - \alpha)$ of including μ.

In cases where a uniform prior distribution is reasonable, a close look at the definition of the posterior distribution (1) leads to the observation that all of the θs on the right side of the equation are in the sampling distribution, $f(Y \mid \theta)$. The denominator is a function of Y alone; θ is integrated out. And if θ has a uniform prior, then $\pi(\theta)$ = constant, so there are no θs there, either. The consequence is that the posterior distribution $f(\theta \mid Y)$ is proportional to $f(Y \mid \theta)$; this latter, when considered as a function of θ, is the likelihood function from which maximum likelihood estimators are obtained. If the posterior distribution is proportional to the likelihood, they both are maximized by the same θ. Thus, in a flat prior Bayesian analysis, the mode of the posterior dis-

tribution is the same as the maximum likelihood estimator (MLE). Consequently, analysts who prefer the Frequentist philosophy to the Bayesian philosophy can still use Bayesian tools for model fitting. In particular, MCMC is a Bayesian tool for evaluating posterior distributions; specifying a flat prior analysis, a Frequentist can use MCMC to find MLE.

MARKOV CHAIN MONTE CARLO

The usefulness of Bayesian methods has been limited by difficulties associated with the calculation of the posterior distribution (1). The normal–normal mixture in the preceding section involves a conjugate prior: similarities in the functional forms of the prior and sampling distributions cause the posterior to be of the same form as the prior, and hence easily calculable. However, for many models, calculation of the integral in (1) is prohibitively difficult; this is especially true for hierarchical models, such as the kittiwake model we describe later. Markov chain Monte Carlo methods provide a solution to this problem.

Markov chain Monte Carlo methods are extensions of the ordinary Monte Carlo (simulation) methods familiar to readers of *The Journal of Wildlife Management* (*JWM*)—151 of 2,113 *JWM* papers indexed in the *BIOSIS* database during 1985–2000 include "Monte Carlo," "bootstrapping," or "simulation" in their abstracts or keyword lists (Biological Abstracts 2001). Monte Carlo methods are used in evaluating model predictions (e.g., in population viability analyses [PVA]). They are used to evaluate the distributions of test statistics, especially when asymptotic approximations are inadequate. Bootstrapping is another familiar Monte Carlo method that is used to evaluate the bias and variability of estimation procedures; bootstrapping is distinguished from other Monte Carlo methods because the simulated values are drawn from an estimated rather than a fully specified model (Manly 1994).

The common feature of Monte Carlo applications is that simulation is used in place of an intractable mathematical calculation. Interest focuses on some function of model parameters $g(\theta)$ that can be expressed as the expected value of a function of data, $h(Y)$. Typically, an analyst simulates independent data sets Y_i^s, $i = 1, 2, \ldots, N$, and approximates $g(\theta)$ by

$$\hat{g}(\theta) = \frac{1}{N} \sum_{i=1}^{N} h(Y_i^s). \qquad (7)$$

In a PVA, for example, $h(Y)$ could be a zero-one indicator of population extinction, and $g(\theta)$ the extinction rate. A Monte Carlo evaluation of the bias of an estimator $T(Y)$ of a parameter $q(\theta)$ is described by (7) with $h(Y) = T(Y) - q(\theta)$.

Given the difficulties associated with calculating posterior distributions, it is natural to question whether some sort of Monte Carlo approach could be applied in a Bayesian analysis. Recall that in a Bayesian analysis, interest focuses on the posterior distribution $f(\theta \mid Y)$; the roles of data Y and parameters θ are switched relative to the sampling distribution $f(Y \mid \theta)$. Thus, a Bayesian Monte Carlo analysis would consist of estimating a function $G(Y)$, expressible as the expected value of a function of parameters $H(\theta)$, with θ sampled from $f(\theta|Y)$. Unfortunately, drawing independent samples from the posterior distribution is not easy due to the difficulties in calculating the integral in (1).

Although it may not be easy to generate independent samples from the posterior, it is easy to produce a first-order Markov chain of values sampled from the posterior distribution, even without calculating the integral in (1). A first-order Markov chain is a sequence of dependent observations, $\{X_t\}_{t=1}^{\infty}$, with the property that the distribution of X_{t+1} given all previous observations, depends only on X_t. Numerous technical details must be considered, but the basic idea is simple: instead of the difficult task of generating a sequence of independent observations from the posterior distribution, the analyst generates a first-order Markov chain of dependent draws θ_i^s from the posterior, and approximates features of the posterior distribution, $G(Y)$, in analogy with (7) by

$$\hat{G}(Y) = \frac{1}{N} \sum_{i=M+1}^{N+M} H(\theta_i^s); \qquad (8)$$

note that, for technical reasons described later, the first M values from the Markov chain are discarded. The mean, median, standard deviation, percentiles, and other descriptors of the posterior distribution can be approximated in this fashion, allowing a Bayesian analysis of the data Y. Non-Bayesian applications of MCMC choose flat priors, and approximate maximum likelihood estimates by the posterior mode.

METROPOLIS-HASTINGS ALGORITHM

In this section, we describe the Metropolis-Hastings algorithm, which is the basis for producing the Markov chains used in MCMC. We begin with a brief review of certain features of Markov chains.

A mental picture of a Markov chain can be made by considering the activities of an absent-minded

and aimless bureaucrat with strong habits. The bureaucrat has no particular goal in mind, wandering from activity to activity, deciding what to do next simply on the basis of what he currently is doing (all of the Markov chains we discuss in this paper are first-order). The set of activities is called a state space, and the changes in activities are transitions. The bureaucrat's absent-mindedness is the defining Markov property. Transition probabilities are entirely determined by his current state, without regard to what he has done previously. The bureaucrat's strong habits make his a stationary Markov chain: not only are his transition probabilities determined by his current state, but they do not evolve through time. If there is a 30% chance that he goes from reading the newspaper to looking out the window on Thursday, the same transition probability will hold on Friday.

Many bureaucrats wander into some subset of activities from which they never emerge, or at least not for arbitrarily long numbers of transitions. If we require that our bureaucrat act otherwise, his Markov chain will be irreducible and positive recurrent. That is, we require that each state can be reached from each other state, and that the average return time to a given state is finite. A feature of stationary, irreducible, positive recurrent Markov chains is that they have a stationary distribution describing the long-term relative frequency of time spent in each state, and that this stationary distribution is attained independent of the initial state of the system. Biologists familiar with age- and stage-structured projection matrices (Caswell 2001) will recognize the property of strong-ergodicity, which leads to stage distributions independent of the initial state (stage structure) of the system.

Suppose that there are 4 possible behavioral states for the bureaucrat (W = looking out the window, N = reading the newspaper, E = checking e-mail, and C = visiting at the water cooler) and that every 10 min his watch beeps to remind him to make a transition. A matrix of transition probabilities may look as follows:

$$\text{From:}\quad \begin{array}{c} \\ W \\ N \\ E \\ C \end{array} \begin{array}{c} \text{To:} \\ \begin{array}{cccc} W & N & E & C \end{array} \\ \begin{bmatrix} 0.45 & 0.48 & 0.07 & 0 \\ 0.30 & 0.20 & 0.10 & 0.40 \\ 0.55 & 0.05 & 0.39 & 0.01 \\ 0.41 & 0 & 0.49 & 0.10 \end{bmatrix} \end{array} \qquad (9)$$

Thus, for instance, if the bureaucrat is reading the newspaper, there is a 30% chance that he will start looking out the window, 20% chance that he will continue reading the newspaper, 10% chance that he will check his e-mail, and a 40% chance that he will hop up to visit the cooler. These probabilities determine his stationary distribution: denoting by P the transition matrix in (9), the stationary distribution can be found either by multiplying P by itself numerous times (the rows will converge to the stationary distribution) or by finding the eigenvector corresponding to the dominant eigenvalue of P transpose: the percentages of time spent in activities W, N, E, and C are specified by a vector V = [42.4%, 26.6%, 18.9%, 12.1%].

The usual question is: "Given transition matrix P, what is the corresponding stationary distribution V?" Markov chain Monte Carlo turns the question around: "Is there a transition matrix P yielding a specified stationary distribution V?" In particular, we seek a transition matrix for a Markov process taking values in the range-space of θ, with stationary distribution equal to the posterior distribution $f(\theta \mid Y)$. If we could produce such a Markov chain, we could use it to make inferences based on the posterior distribution. The mean value of the chain would converge to the Bayes estimate of θ. Given flat priors, the posterior mode of the chain would approximate the maximum likelihood estimator of θ. The central 95% range of values of the chain would approximate the central 95% region of the posterior distribution, creating a Bayesian confidence interval for θ. In short, if we could produce such a Markov chain, it could be used as the basis of a Markov chain Monte Carlo calculation, as described by (8).

It turns out that this is quite easily accomplished. A simple algorithm, due to Metropolis et al. (1953) and Hastings (1970) does the job. Its transition structure is as follows: from state θ_t, generate a candidate state θ^* by random sampling from a distribution $C(\theta^* \mid \theta_t)$. Next, compute

$$r(\theta_t) = \min \left\{ \frac{f(\theta^* \mid Y)/C(\theta^* \mid \theta_t)}{f(\theta_t \mid Y)/C(\theta_t \mid \theta^*)}, 1 \right\}, \qquad (10)$$

and determine the next state of the chain on the basis of a Bernoulli trial:

$$\theta_{t+1} = \begin{cases} \theta^* \text{ with probability } r(\theta_t) \\ \theta_t \text{ with probability } 1 - r(\theta_t) \end{cases}; \qquad (11)$$

the process stays where it is with probability $1 - r(\theta_t)$, or moves to the candidate value with probability $r(\theta_t)$. The resulting sequence θ_t, $t = 1, 2, \ldots$, is a Markov chain with the desired stationary distrib-

ution. For a heuristic description of why the algorithm works see Gelman et al. (1998:325). The performance of the algorithm depends on the choice of candidate generating distribution, $C(\theta^*|\theta_t)$, as discussed subsequently, and may sometimes be improved by multiplying the transition probability by a scalar α, $0 < \alpha < 1$ (Goggins et al. 1998).

The movement probabilities (10) depend on the posterior distribution only through ratios of values at the present and candidate values. The beauty of this feature is that the integral in (1), which sometimes stands in the way of calculating the posterior, cancels out. Thus, MCMC can be used to approximate features of the posterior distribution, even when the posterior distribution itself cannot be calculated.

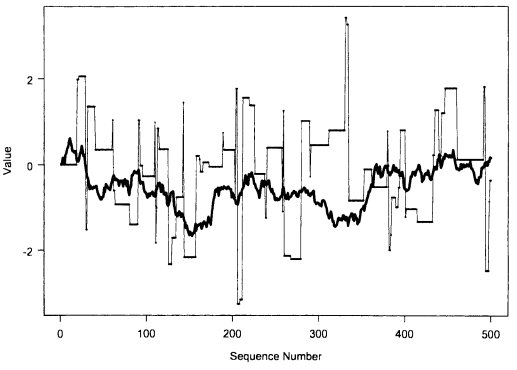


Fig. 1. Realizations of 2 Markov chains produced using the Metropolis-Hastings algorithm, each with standard normal stationary distribution. Candidate values produced by adding mean 0 normal noise to current values. Standard deviations of candidate values were 0.10 (thick curve), and 10.0 (thin curve).

IMPLEMENTATION OF MCMC

Markov chain Monte Carlo is becoming a standard technique in the repertoire of applied statisticians. A significant contribution to the field has been the development and distribution of program *BUGS* (Spiegelhalter et al. 1995). Gibbs sampling is a particularly adroit implementation of the Metropolis-Hastings algorithm. (For details on Gibbs sampling see Casella and George [1992]; for its interpretation as a special case of the Metropolis-Hastings algorithm, see Gelman et al. [1998:328].)

With the availability of software such as *BUGS* and the increasing speed of personal computers, we anticipate that MCMC will be routinely applied to biological models. Users of MCMC need not have a sophisticated knowledge of the details of its implementation—MCMC is simply a procedure for simulation. A practitioner needs to be aware of 2 important issues for evaluating the performance of an MCMC simulation. These relate to the distinguishing feature of MCMC simulation—rather than drawing independent samples from a target distribution, MCMC produces a Markov chain of values with the target distribution as its stationary distribution. This distinction has important consequences for the precision and accuracy of simulation summaries.

Precision and Autocorrelation

Most summaries of simulated data are averages of 1 sort or another (as equations [7] and [8]). The precision of averages increases with the number of observations (i.e., the number of simulations) but is reduced by positive correlation, such as in a Markov chain. For chains generated using the Metropolis-Hastings algorithm, the magnitude of the correlation depends on the relation between current values θ_t and candidate values θ^*. If candidate values are too close to current values, sampling of the target distribution will be slow, and autocorrelation will be high. On the other hand, if θ^* is likely to be too far from θ_t, it may not represent the posterior distribution and consequently will have a small acceptance probability (10). In this case, the candidate values tend to be rejected, the chain does not move (i.e., $\theta_{t+1} = \theta_t$), and autocorrelation will be high.

The 2 extremes are illustrated in Fig. 1. We generated 2 Markov chains using the Metropolis-Hastings algorithm. We obtained candidate values by adding normally distributed noise to the current values. Both chains have standard normal stationary distributions. The lighter line corresponds to a candidate distribution with standard deviation of 10—83% of the candidate values were rejected, and the lag-1 autocorrelation was 0.86. The heavier curve was generated using a candidate distribution with standard deviation of 0.10—only 4% of the candidate values were rejected, but the step sizes were very small, and the lag-1 autocorrelation was 0.97.

It is possible to tune the choice of candidate distribution to optimize the performance of the MCMC simulation. For the example presented here, Fig. 2 presents empirical evidence that setting the standard deviation of the candidate distribution to an intermediate value of 2.3 minimizes the lag-1 autocorrelation at a value of about 0.63. Such matters may be beyond the concern or interest of most users of MCMC. However, users

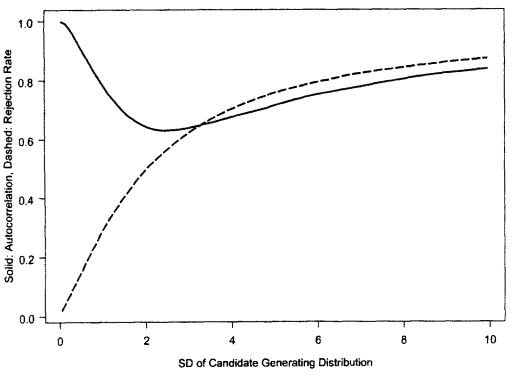


Fig. 2. Autocorrelation (solid line) and rejection rate (dashed line) for Metropolis-Hastings Markov chains with standard normal stationary distributions and candidate values generated by adding mean 0 normal noise to current value, plotted as a function of standard deviation of candidate generating distribution.

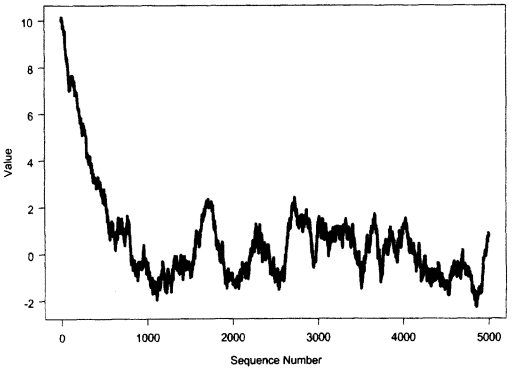


Fig. 3. Transient behavior of Metropolis-Hastings Markov chain resulting from the starting value outside of range of stationary distribution. The stationary distribution of the chain is standard normal. Candidate values were produced by adding mean 0 normal noise with standard deviation of 0.10 to current values.

should be aware of the existence of problems relating to excessive autocorrelation and routinely evaluate the results of MCMC simulations for its effects.

Accuracy and Starting Values

Another consideration in summaries of MCMC relates to accuracy—whether the chain of values sampled can be thought of as a sample from the target distribution. At issue is the starting value of the chain, which must be supplied by the practitioner. If the starting value could be sampled from the target distribution there would be no problem: every value in the Markov chain could be considered a sample from the target distribution. However, if the starting value does not represent the target distribution, neither will other early values in the chain; the chain exhibits transient behavior, moving gradually into the range of the target distribution.

This phenomenon is clearly evident in Fig. 3. We generated the Markov chain according to the Metropolis-Hastings algorithm, with a standard normal target distribution. We obtained candidate values by adding normally distributed noise with standard deviation of 0.10 to the present values. The starting value of 10 being well outside the range of the stationary distribution, the chain exhibits transient behavior—a burn-in period of 500 or 1,000 values is needed before the chain appears to be sampling in the range of the stationary distribution. Thus, in equation (8), an analyst would set $M = 1{,}000$, and discard these values from MCMC calculations.

The selection of candidate distributions and assessment of convergence to the stationary distribution are areas of active research. For a more in-depth, yet readable account of the issues involved, readers are referred to the report of a roundtable discussion of MCMC in practice (Kass et al. 1998).

The value of MCMC as a statistical tool becomes apparent when one considers fitting complex statistical models to data. In the next section, we illustrate the use of *BUGS* to analyze a moderately complex hierarchical model, one that defies analysis by conventional methods.

AN ILLUSTRATION OF MCMC USING *BUGS*

J.-Y. Monnat, E. Danchin, and a team of collaborators have studied several colonies of kittiwakes on the Brittany coast of France for 20 years (Danchin and Monnat 1992, Danchin et al. 1998). Here, we describe an analysis of association in individual-specific latent factors governing survival and breeding rates. The data set consists of survival and breeding records over 13 years for 845 birds known to have bred at least once. The question of interest was whether trade-offs exist between components of fitness. For example, do birds that are more productive tend to have lower survival rates (Stearns 1992, Cam et al. 1998)?

The collection of data is described elsewhere (Cam et al. 2002), but several important features are noted here. First, we note that new birds of known age were individually marked through the study period; the number of individuals first included in years 1, 2, ... ,12 were 98, 82, 88, 97, 76, 30, 73, 69, 54, 52, 60, and 66, respectively. This

feature of the data allows the analyst to distinguish age and year effects on breeding and survival.

Another important feature of these data is that the resighting probability for marked birds was 1 (Cam et al. 1998). Thus, for each bird alive in year t, there is a record of whether the bird attempted breeding in year t, and of whether it survived to year $t+1$. We modeled the 2,947 survival events, S, as Bernoulli trials with success parameters, φ, and the 2,344 breeding events, B, as Bernoulli trials with success parameter, β. The effects of covariates on survival and breeding probabilities were modeled as linear on the logit scale, $\text{logit}(p) = \ln(p/(1-p))$.

We evaluated a cross-classification of 32 candidate models. All of the models included year effects on breeding and survival as factors. We modeled age effects on survival and breeding as an additive factor, a quadratic function, a linear function, or constant. We compared the adequacy of these models with and without bird-specific effects. The model selection process is reported elsewhere (Cam et al. 2002), along with a discussion of the biological implications of the model. Here, we describe the fitting of the model that was eventually selected using Akaike's Information Criterion.

The selected model included bird-specific effects, specified that the effect of aging on survival was linear, and that the effect of aging on breeding was quadratic. Thus, the model specified that

$$\text{logit}(\varphi_{bird,yr}) = \mu_\varphi + a_\varphi A_{bird,yr} + \delta^\varphi_{yr}, + \alpha^\varphi_{bird},$$

and (12)

$$\text{logit}(\beta_{bird,yr}) = \mu_\beta + \alpha_\beta A_{bird,yr} + b_\beta A^2_{bird,yr} + \delta^\beta_{yr} + \alpha^\beta_{bird};$$

here, A is age, δ is year effect (a factor having 12 levels each, the first of which was set equal to 0 for identifiability), and α is individual effect. Sub- and superscripts φ and β distinguish parameters related to survival and breeding, respectively; subscripts *bird* and *yr* identify individual birds, *bird* = 1, 2, ..., 845, and years, *yr* = 1, 2, ..., 12. Pairs of bird-specific parameters, $(\alpha^\varphi_{bird}, \alpha^\beta_{bird})$ are of special interest in this analysis.

Clearly, it is neither practical nor desirable to include 1,690 = 845 × 2 parameters describing individual effects in the model. Even were such an approach feasible, the analyst would end up calculating statistics on statistics, summarizing and evaluating the collection of parameter estimates, attempting to correctly account for the sampling variation in the estimates. Instead, we treated individual effects as bivariate random effects. Thus, corresponding to each bird is an unobservable pair of latent effects, a bivariate parameter $\alpha_{bird} = (\alpha^\varphi_{bird}, \alpha^\beta_{bird})'$. These are assumed to have bivariate normal distributions, with mean 0, and variance–covariance matrix Σ. Therefore, the individual effects are described by 3 parameters: 2 variances and a correlation. The correlation in individual effects is of special biological interest. Negative values would indicate that birds with higher survival probabilities are less likely to attempt breeding, given that they survive. Positive correlations indicate that birds with higher survival probabilities are more likely to attempt breeding, given that they survive.

Calculation of the likelihood for this model is prohibitively difficult, so fitting by conventional methods is not possible. However, an objective Bayes analysis is fairly straightforward using *BUGS*.

Analysis in *BUGS* is aided by the specification of a directed acyclic graph as illustrated in Fig. 4, and explained below. A directed acyclic graph is a visual metaphor for a hierarchical model, consisting of nodes, plates, and edges (respectively, the ellipses, rectangles, and arrows, in Fig. 4). Nodes are drawn for each parameter and for the data, and for parameters that are obtained as functions of other parameters. Arrows are added specifying hierarchical relations and dependencies; collections of related nodes are summarized by plates. Thus, in Fig. 4, the lower left-hand plate has nodes S[i] for the 2,947 survival events in the data set; these are modeled as Bernoulli trials, with success rates phi[i], which is a function of the quantities represented by nodes with arrows leading to it.

Highlighting a node in *BUGS* allows the specification of whether the node is stochastic, logical (i.e., a deterministic function of the values of other nodes), or constant. In Fig. 4, B[J] (the jth breeding event) is seen to be stochastic, a Bernoulli trial with success parameter beta[J]. If the node beta[J] were highlighted, it would be observed to be a logical node, calculated by the specification

```
logit(beta[j]) = mu_beta + a_beta *
(BreedAge[j]) + b_beta * pow(BreedAge[j],2)
        + delta_beta[BreedYr[j]]
        + alpha_beta[BreedID[j]],        (13)
```

corresponding to the second portion of the model specification (12).

Founder nodes, those without arrows leading to them in the directed acyclic graph, must either

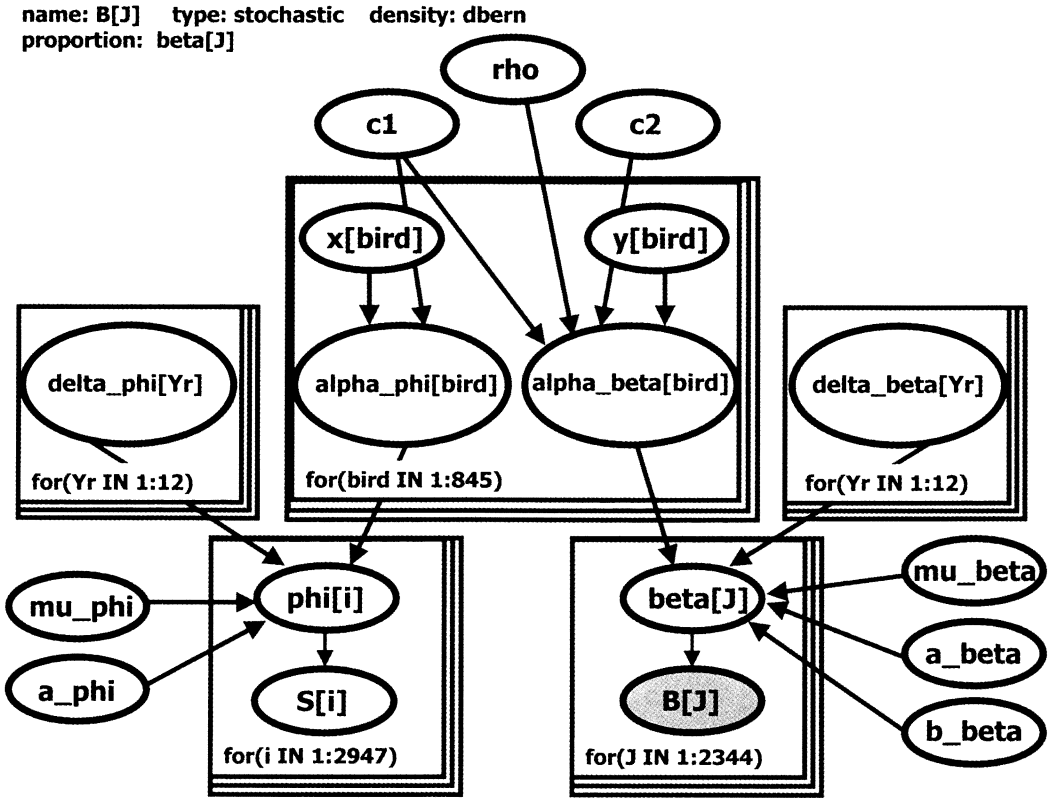


Fig. 4. Directed acyclic graph for kittiwake model (equation [12]), from program *BUGS*. Nodes are drawn for each parameter and for the data, with arrows specifying hierarchical relations and dependencies. Highlighted node B[J] indicates a stochastic type, with density *dbern* and parameter *beta*[*J*]. The node represents a Bernoulli trial with success parameter specified by node *beta*[*J*].

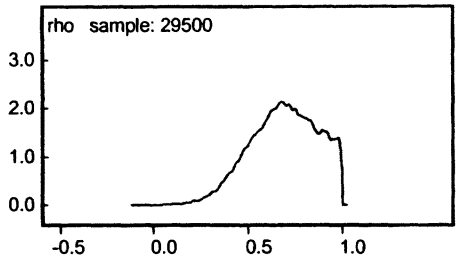
be pre-specified constants or random variables. Thus in Fig. 4, the nodes corresponding to μ_φ, a_φ, μ_β, a_β, b_β, δ^φ_{yr}, and δ^β_{yr} in the model specification (12) are random variables (as in a Bayesian analysis) rather than fixed but unknown constants (as in a Frequentist analysis). Since we desired an objective Bayes analysis, we specified prior distributions that were mean zero normal with standard deviations of 1,000. These priors are essentially uniform over a large range. The density values on [–50,50] are always within 99.87% of the maximum value. Values outside of that range are meaningless on the logit scale. Therefore, we may consider these priors as essentially uniform. As previously mentioned, it follows from equation (1) that the maximizer for the posterior distribution of θ is the same as the maximum likelihood estimator. Thus, the modes of the posterior distributions for μ_φ, a_φ, μ_β, a_β, b_β, δ^φ_{yr}, and δ^β_{yr} will be the maximum likelihood estimators in a Frequentist analysis.

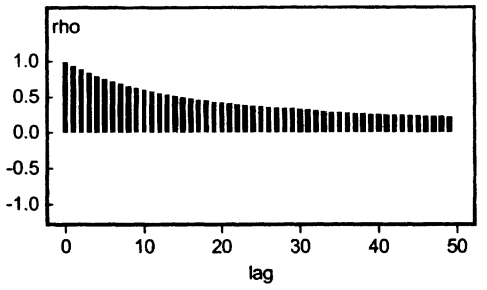
The remaining founder nodes in Fig. 4 relate to the latent individual-specific effects. Nodes X[bird] and Y[bird] are independent standard normal variates, linear combinations of which were calculated so as to produce pairs $\alpha_{bird} = (\alpha^\varphi_{bird}, \alpha^\beta_{bird})'$ with covariance matrix

$$\Sigma = \begin{bmatrix} c_1^2 & \rho c_1 c_2 \\ \rho c_1 c_2 & c_2^2 \end{bmatrix}. \qquad (14)$$

We specified vague inverse gamma priors (a standard noninformative prior for variances) for the 2 variances, and gave ρ a uniform prior on [–1, 1].

Once the model has been completely specified, *BUGS* selects an appropriate algorithm and produces Markov chains for each of the parameters. These chains can be output as ASCII files for analysis in other software packages (e.g., a collection of S-Plus functions named CODA is available from the *BUGS* website). Some graphical and descriptive summaries also are available within *BUGS*. Some





node	mean	sd	MC error	2.5%	median	97.5%	start	sample
rho	0.6936	0.1769	0.007207	0.3323	0.6984	0.9825	501	29500

Fig. 5. Summaries from Program *BUGS* of a Markov chain of values of ρ from analysis of model (12) include a smoothed histogram, an estimated autocorrelation function, and a tabular summary. Tabular summary provides approximate values for the mean, standard deviation, and percentiles 2.5, 50, and 97.5 of the posterior distribution. These were computed from a chain of length 30,000, the first 500 values of which were discarded as a burn-in period.

of these are shown in Fig. 5, for the parameter ρ.

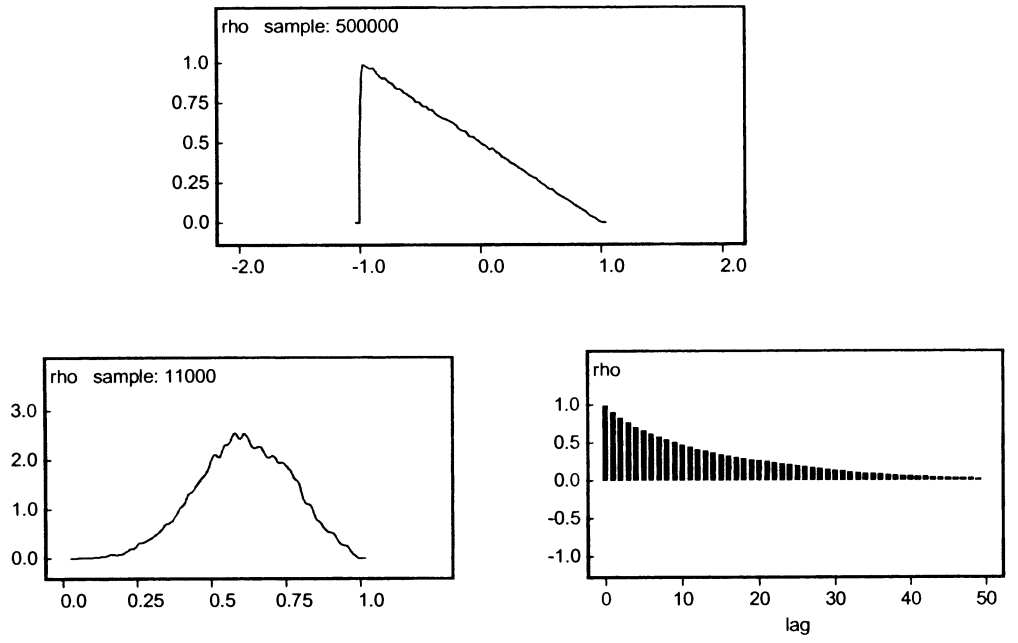
Fig. 5 displays a smoothed histogram and an estimate of the autocorrelation function of the Markov chain of values of ρ. The tabular summary for ρ shows approximate values for the mean, standard deviation, and percentiles 2.5, 50, and 97.5 of the posterior distribution; also, that these were computed from a Markov chain of length 30,000, the first 500 values of which were discarded as representing a burn-in period.

The interval (0.33, 0.98) contains the central 95% of the mass of the posterior distribution and sometimes is called a 95% Bayesian confidence interval or credible interval, although these terms sometimes are reserved for the shortest interval containing 95% of the posterior distribution. In the present case, the shortest such interval is (0.39, 1]. The mean and median values of the posterior distribution are both close to the mode (estimated, outside of *BUGS*, as 0.68 using a histogram smoother).

As mentioned at the outset, the possibility of varying the prior is an asset, rather than a liability of Bayesian analysis. In Fig. 6, we report the results of an analysis with an informative prior (triangular distribution at top). This prior places 3:1 odds on $\rho < 0$, and 15:1 odds on $\rho < 0.50$, and hence is hardly objective. This prior might represent strongly held views of individual trade-offs in breeding and survival probabilities. Under this prior, the 95% credible interval [0.28, 0.90] and posterior mean (0.61) are only slightly changed from those obtained using the uniform prior ([0.33, 0.98] and 0.69). This sort of result often is described as the data overwhelming the prior—the contribution of Prior to Posterior is small relative to that of Data, in equation (2).

The evidence is that $\rho > 0$: latent factors governing survival rates and breeding rates are positively correlated. Thus, the birds that are more likely to survive also are more likely to breed, given that they survive. There is no evidence of a trade-off at the individual level.

We discuss some of the biological and management implications of this finding in the next section; a more detailed discussion is in Cam et al. (2002). The 2 main conclusions are: (1) differences in survival and breeding probabilities among individuals are substantial; models including individual effects systematically had a better fit than others, which supports the hypothesis of heterogeneity in vital rates among individuals; and (2) the pattern of age-related variation in breeding and survival rates detected at the individual level differed from that observed at the population level. Our approach provided evidence of senescent decline in survival. Such a decline was undetectable when analyzed using classic approaches to the effect of age on survival. This phenomenon has been extensively studied in humans (Vaupel and Yashin 1985*a,b*) and captive animals (Service 2000), but very rarely in wild animal populations (McDonald et al. 1996). From a modeling perspective, the analysis includes estimation of variation and covariation in fitness components at the individual level. Evaluation of these fitness components is crucial for the study



node	mean	sd	MC error	2.5%	median	97.5%	start	sample
rho	0.6072	0.1575	0.007869	0.2835	0.6076	0.8966	501	11000

Fig. 6. Summaries from Program *BUGS* for an alternative analysis of model (12). Top panel is a histogram of 500,000 draws from an informative prior representing strong prior convictions that $\rho < 0$. Histogram, estimated autocorrelation, and tabular summaries of posterior distribution based on a chain of values of 11,000 after a burn-in of 500.

of age-specific reproductive strategies (Charlesworth 1994), and is also important for studies of natural selection (Endler 1986). Accurate description of age effects has major implications for designing management plans as well (see below). Markov chain Monte Carlo allowed the fitting of what otherwise might have been regarded as prohibitively complex models.

THEORETICAL AND MANAGEMENT IMPLICATIONS OF INDIVIDUAL HETEROGENEITY

Individual Variation and Population Dynamics

Understanding the processes underlying population dynamics is of fundamental interest in most conservation programs, or to design management plans (Nichols et al. 2000). Many classical ecological models are based on the assumption that populations consist of identical individuals. However, as emphasized by Bjørnstad and Hansen (1999), extensive evidence exists that natural populations exhibit much genetic and nongenetic variation in life-history traits and demographic parameters. The sources of variation among individuals coexisting in a population at a given point in time most commonly incorporated into population models are age, size, or stage (i.e., Caswell 2001). However, our results show that these sources of variation may not be sufficient to adequately account for heterogeneity among individuals.

The population consequences of variation among individuals in vital rates associated with differences in access to resources resulting from social hierarchy or habitat heterogeneity, sometimes combined with stage or age, are receiving growing interest (e.g., de Jong 1979, Hassell and May 1985, Łomnicki 1988, Bjørnstad and Hansen 1999, Bjørnstad et al. 1999). Results obtained using models incorporating these sources of heterogeneity indicate that individual variation can influence population growth rate, equilibrium density, and stability (Bjørnstad and Hansen 1999). Doebelli and de Jong (1999) investigated

the influence of genetic variability in sensitivity to population density and reached the same conclusions. They extended their conclusions to environmentally induced variability. Other sources of variation among individuals and their consequence on population dynamics also are receiving growing interest, such as cohort and maternal effects (Hansson 1984, Albon et al. 1987, Schluter and Gustafsson 1993, Sedinger et al. 1995, Boonstra and Hochachka 1997, Rose et al. 1998, Lindström 1999). Both sources of heterogeneity among individuals are suspected to have lagged demographic consequences and result in destabilization of the population (Albon et al. 1992, Ginzburg 1998). However, Bjørnstad and Hansen (1999) also emphasized that the consequences of individual heterogeneity strongly depends on the form of the model, and further work is needed to reach generally applicable conclusions. So far, attempts to use age-structured models accounting for additional sources of individual heterogeneity in vital rates have been rare, mostly because of the complexity of the models required (Lomnicki 1988).

The importance of individual variation recently has been emphasized in studies focusing on the relationship between individual decisions and population dynamics (e.g., Marrow et al. 1996). Extensive evidence exists that the number of young produced in a population during a given year is influenced by environmental conditions (e.g., resource availability). This type of effect could be viewed as a simple constraint shaping the reproductive potential of individuals at that point in time only. However, a fundamental assumption of life-history theory is that individual reproductive decisions depend not only on the specific conditions at that point in time and their consequences on the probability of raising young successfully, but also on future fitness (i.e., trade-offs between current and future fitness; Stearns 1992). Recent developments in that field (e.g., McNamara and Houston 1992, 1996; Clark and Mangel 2000) rely on the influence of individual state (e.g., condition) on decisions, individuals in different states making different decisions in terms of age of first breeding, clutch or litter size, migration routes, or dispersal (e.g., Festa-Bianchet and Jorgenson 1998). Individual state reflects both underlying differences between individuals (i.e., the type of differences addressed in this study) and the influence of environmental conditions recently experienced (McNamara and Houston 1992, 1996). As the current and future fitness prospects of individuals with different underlying survival or breeding potential are not the same, individuals are expected to make different decisions in the same environmental conditions. The overall production of young in a population during a given year may depend on the proportion of individuals with different underlying characteristics. Assessing the distribution of individuals with different underlying vital rates is critical to understanding the population consequences of individual decisions.

In addition to the difficulty raised by the complexity of population models accounting for many sources of individual variation in demographic parameters, quantifying this heterogeneity in wild animal populations is a challenge (Cooch et al. 2002). The models used here to assess individual variation in breeding and survival rates can be described as models where each individual has its own mortality risk (Service 2000), or its own reproductive potential. In this study, we accounted for the correlation between breeding and survival rates at the individual level; fitting such models using conventional approaches is prohibitively difficult (Cam et al. 2002). One of the strategies to build population models accounting for individual variation is to assign a growth rate depending on some parameter with an individual value to each individual in the population (Bjørnstad and Hansen 1999). Now, as emphasized by Bjørnstad and Hansen (1999), the influence of individual variation on population dynamics is strongly dependent on the distribution of the parameter in the population. Describing the form of variation among individuals thus is critical to understanding the consequences of individual heterogeneity on the dynamics of the population. Development of approaches to fitting complex hierarchical models permitting estimation of the distribution of the parameters in the population is a promising advance and should provide means of addressing the population consequences of individual variation in life-history traits and demographic parameters using empirical data.

Individual Variation: Management and Conservation

The potential relevance of individual heterogeneity to harvest management was explicitly noted by Johnson et al. (1986, 1988). They considered the hypotheses of additive and compensatory hunting mortality as originally described for waterfowl by Anderson and Burnham (1976). They then considered a heterogeneous population in which probabilities associated with both hunting and nonhunting mortality differed

among individuals and were positively associated within individuals (individuals with low nonhunting mortality probabilities also had low hunting mortality probabilities). Even in the case where hunting mortalities acted as instantaneous competing risks (additive hunting mortality hypothesis), the population appears to compensate for increased harvest mortality. The mechanism underlying this compensation does not involve density-dependent changes in nonhunting mortality, but simply results from a postharvest population that contains an increased proportion of low-mortality individuals. Patterns of individual heterogeneity that showed no covariance within individuals for hunting and nonhunting mortality probabilities, or that showed a negative covariance, would yield different responses to harvest. In fact, a negative covariance would yield an especially severe population-dynamic effect of harvest. In addition to the effect of harvest mortality, the postharvest population would have a higher nonhunting mortality probability than would a nonharvested population. The point is that individual heterogeneity in vital rates, specifically the patterns of variation and covariation of vital rates among individuals, are very relevant to population responses to management actions such as harvest.

The concept of reproductive value is important to various kinds of management and conservation problems. With respect to conservation, the age classes with the largest reproductive value make the largest contributions to future population growth and are thus selected for restocking and reintroduction programs (see MacArthur and Wilson 1967). With respect to harvest management, if all age classes are of equal value to the hunter, then the age classes with the smallest reproductive values should be harvested (e.g., MacArthur 1960). The shape of the reproductive value function is heavily dependent on whether or not individuals experience senescent decline in either survival or reproductive rates (e.g., see Nichols et al. 1980, Fig. 4). If our perception of senescence is obscured by individual heterogeneity (as illustrated in the kittiwake example above; see also Cam et al. 2002), then management actions that are based on reproductive value or—more generally—age-specific patterns of survival and reproduction, may be misdirected. It is important to draw strong inferences about the existence of senescent decline in vital rates among individuals. We do not know how to draw such inferences in the absence of analytic approaches that properly incorporate individual heterogeneity in vital rates.

Population viability analysis (PVA) has become a popular tool in certain areas of conservation biology. Population viability models frequently include temporal and sometimes spatial variation in vital rates, but typically have not included individual heterogeneity in birth and death rates. However, such heterogeneity can be very important for estimating extinction probabilities associated with populations. Specifically, extinction probabilities have been shown to typically decrease in the presence of individual heterogeneity in vital rates (Conner and White 1999). Results of models that do, and do not, incorporate individual heterogeneity can show substantial differences in such quantities as persistence time and probability of extinction (Conner and White 1999, White 2000).

The above examples illustrate the importance of individual heterogeneity to population management and conservation. The relevance of individual heterogeneity in vital rates extends well beyond these examples to virtually all management or conservation actions. Patterns of variation and covariation of vital rates among individuals are important determinants of population dynamics and responses to management. We believe that the usefulness of the methods presented here for estimating such variation and covariation will be increasingly recognized in wildlife management and conservation biology.

DISCUSSION

Markov chain Monte Carlo is a tool for fitting complex statistical models to data. It is becoming more widely used in a variety of biological and related fields (Fig. 7). We found 235 citations including "Gibbs sampling," "Markov chain Monte Carlo," "Metropolis-Hastings," or related terms in *Biological Abstracts* (2001). Most of the citations were in statistics journals (83×: including *Statistics in Medicine* 30×, *Biometrics* 26×, and *Biometrika* 9×); genetics journals (63×: including *Genetics, Selection, Evolution* 18×, *Genetic Epidemiology* 13×, and *American Journal of Human Genetics* 8×); and animal science journals (36×: including *Journal of Dairy Science* 15× and *Journal of Animal Science* 9×). None of these terms were found in *The Journal of Wildlife Management, Ecological Applications,* or *Ecology.*

This is not to say that none of these applications would interest readers of *The Journal of Wildlife Management.* In particular, we note extensions of conventional mark–recapture methodologies (George and Robert 1992, Dupuis 1995, Vounatsou and Smith 1995), and of the Cox proportional hazards model (Gauderman and

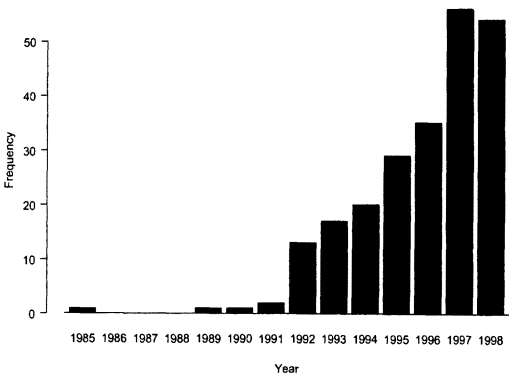


Fig. 7. Number of citations found in *BIOSIS* database for 1985–1998 including "Gibbs sampling," "Markov chain Monte Carlo," "Metropolis-Hastings" or related terms in their abstracts or keyword lists; plotted against publication year.

Thomas 1994, Goggins et al. 1998, Sargent 1998) using MCMC. Applications of MCMC in ecological journals include autologistic modeling for the spatial distribution of wildlife (Augustin et al. 1996), modeling of nitrogen flows in the world's oceans (Harmon and Challenor 1997), and assessment of fisheries stock (Meyer and Millar 1999).

Given the growing number of applications of MCMC in all branches of statistical application, the availability of software such as *BUGS*, the spectacular rate of increase in clock speeds of personal computers, the consequent feasibility of ever more computationally intensive statistical methods, and the potential for fitting complex hierarchical models, we predict MCMC will be used as a tool in many future wildlife applications.

LITERATURE CITED

ALBON, S. D., T. H. CLUTTON-BROCK, AND F. E. GUINNESS. 1987. Early development and population dynamics in red deer. II. Density-independent effects and cohort variation. Journal of Animal Ecology 56:69–81.

———, ———, AND R. LANGVATN. 1992. Cohort variation in reproduction and survival: implications for population demography. Pages 15–21 *in* R. D. Brown, editor. The biology of deer. Springer-Verlag, New York, USA.

ANDERSON, D. R., AND K. P. BURNHAM. 1976. Population ecology of the mallard. VI: The effect of exploitation on survival. U.S. Fish and Wildlife Service Resource Publication 128.

AUGUSTIN, N. H., M. A. MUGGLESTON, AND S. T. BUCKLAND. 1996. An autologistic model for the spatial distribution of wildlife. Journal of Applied Ecology 33:339–347.

BIOLOGICAL ABSTRACTS. 2001. BIOSIS Previews, File 5 [electronic database]. The Dialog Corporation, Cary, North Carolina, USA.

BJØRNSTAD, O. N., J.-M. FROMENTIN, N. C. STENSETH, AND J. GJØSÆTER. 1999. Cycles and trends in cod populations. Proceedings of the National Academy of Sciences of the United States of America 96:5066–5071.

———, AND T. F. HANSEN. 1999. Individual variation and population dynamics. Oikos 69:167–171.

BOONSTRA, R., AND W. M. HOCHACHKA. 1997. Maternal effects and additive inheritance in the collared lemming *Discrostonyx groenlandicus*. Evolutionary Ecology 11:169–182.

CAM, E., J. E. HINES, J.-Y. MONNAT, J. D. NICHOLS, AND E. DANCHIN. 1998. Are adult nonbreeders prudent parents? The kittiwake model. Ecology 79:2917–2930.

———, W. A. LINK, E. G. COOCH, J.-Y. MONNAT, AND E. DANCHIN. 2002. Individual covariation in life-history traits: seeing the trees despite the forest. American Naturalist 159:96–105.

CASELLA, G., AND E. I. GEORGE. 1992. Explaining the Gibbs sampler. American Statistician 46:167–174.

CASWELL, H. 2001. Matrix population models. Second edition. Sinauer Associates, Sunderland, Massachusetts, USA.

CHARLESWORTH, B. 1994. Evolution in age-structured populations. Cambridge University Press, Cambridge, United Kingdom.

CLARK, C. W., AND M. MANGEL. 2000. Dynamic state variable models in ecology. Oxford University Press, New York, USA.

CONNER, M. M., AND G. C. WHITE. 1999. Effects of individual heterogeneity in estimating the persistence of small populations. Natural Resource Modeling 12:109–127.

COOCH, E. G., E. CAM, AND W. A. LINK. 2002. Occam's Shadow: levels of analysis in evolutionary ecology—where to next? Journal of Applied Statistics 29:19–48.

DANCHIN, E., T. BOULINIER, AND M. MASSOT. 1998. Conspecific reproductive success and breeding habitat selection: implications for the evolution of coloniality. Ecology 79:2415–2428.

———, AND J.-Y. MONNAT. 1992. Population dynamics modelling of two neighbouring kittiwake (*Rissa tridactyla*) colonies. Ardea 80:171–180.

DE JONG, G. 1979. The influence of the distribution of juveniles over patches of food on the dynamics of a population. Netherlands Journal of Zoology 29:33–51.

DOEBELLI, M., AND G. DE JONG. 1999. Genetic variability in sensitivity to population density affects the dynamics of simple ecological models. Theoretical Population Biology 56:37–52.

DUPUIS, J. A. 1995. Bayesian estimation of movement and survival probabilities from capture–recapture data. Biometrika 82:761–772.

ENDLER, J. A. 1986. Natural selection in the wild. Princeton University Press, New Jersey, USA.

FESTA-BIANCHET, M., AND J. T. JORGENSON. 1998. Selfish mothers: reproductive expenditure and resources availability in bighorn ewes. Behavioral Ecology 9:144–150.

GAUDERMAN, W. J., AND D. C. THOMAS. 1994. Censored survival models for genetic epidemiology: a Gibbs sampling approach. Genetic Epidemiology 11:171–188.

GELMAN, A., J. B. CARLIN, H. S. STERN, AND D. B. RUBIN. 1998. Bayesian data analysis. Chapman & Hall/CRC Press, Boca Raton, Florida, USA.

GEMAN, S., AND D. GEMAN. 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 6:721–741.

GEORGE, E. I., AND C. P. ROBERT. 1992. Capture–recapture estimation via Gibbs sampling. Biometrika 79:677–683.

GILKS, W. R., S. RICHARDSON, AND D. J. SPIEGELHALTER. 1996. Introducing Markov chain Monte Carlo. Pages 1–20 *in* W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. Markov chain Monte Carlo methods in practice. Chapman & Hall, New York, USA.

GINZBURG, L. 1998. Inertial growth: population dynamics based on maternal effects. Pages 42–53 *in* T. A. Mousseau and C. W. Fox, editors. Maternal effects as adaptations. Oxford University Press, New York, USA.

GOGGINS, W. B., D. M. FINKELSTEIN, D. A. SCHOENFELD, AND A. M. ZASLAVSKY. 1998. A Markov chain Monte Carlo EM algorithm for analyzing interval-censored data under the Cox proportional hazards model. Biometrics 54:1498–1507.

HANSSON, L. 1984. Composition of cyclic and non-cyclic vole populations: on the causes of variation in individual quality among *Clethrionomys glareolus* in Sweden. Oecologia 63:199–206.

HARMON, R., AND P. CHALLENOR. 1997. A Markov chain Monte Carlo method for estimation and assimilation into models. Ecological Modelling 101:41–59.

HASSELL, M. P., AND R. M. MAY. 1985. Individual behaviour and population dynamics. Pages 3–32 *in* R. M. Sibly and R. H. Smith, editors. Behavioural ecology: ecological consequences of adaptive behaviour. Blackwell Scientific, Oxford, United Kingdom.

HASTINGS, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109.

JOHNSON, D. H., K. P. BURNHAM, AND J. D. NICHOLS. 1986. The role of heterogeneity in animal population dynamics. Proceedings of the International Biometric Conference 13:5.3.1–5.3.15.

———, J. D. NICHOLS, M. J. CONROY, AND L. M. COWARDIN. 1988. Some considerations in modeling the mallard life cycle. Pages 9–20 *in* M. W. Weller, editor. Waterfowl in winter. University of Minnesota Press, Minneapolis, USA .

KASS, R. E., B. P. CARLIN, A. GELMAN, AND R. M. NEAL. 1998. Markov chain Monte Carlo in practice: a round-table discussion. American Statistician 52:93–100.

LINDSTRÖM, J. 1999. Early development and fitness in birds and mammals. Trends in Ecology & Evolution 14:343–348.

LINK, W. A. 1999. Modeling pattern in collections of parameters. Journal of Wildlife Management 63:1017–1027.

ŁOMNICKI, A. 1988. Population ecology of individuals. Princeton University Press, New Jersey, USA.

MACARTHUR, R. H. 1960. On the relation between reproductive value and optimal predation. Proceedings of the National Academy of Sciences of the United States of America 46:143–145.

———, AND E. O. WILSON. 1967. The theory of island biogeography. Princeton University Press, New Jersey, USA.

MANLY, B. F. J. 1994. Randomization and Monte Carlo methods in biology. Chapman & Hall, New York, USA.

MARROW, P., J. M. MCNAMARA, A. I. HOUSTON, I. R. STEVENSON, AND T. H. CLUTTON-BROCK. 1996. State-dependent life history evolution in Soay sheep: dynamic modelling of reproductive scheduling. Philosophical Transactions of the Royal Society of London Series B 351:17–32.

MCDONALD, D. B., J. W. FITZPATRICK, AND G. E. WOOLFENDEN. 1996. Actuarial senescence and demographic heterogeneity in the Florida scrub jay. Ecology 77:2373–2381.

MCNAMARA, J. M., AND A. I. HOUSTON. 1992. State-dependant life-history theory and its implication for optimal clutch size. Evolutionary Ecology 6:170–185.

———, AND ———. 1996. State-dependant life histories. Nature 380:215–221.

METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER, AND E. TELLER. 1953. Equations of state calculations by fast computing machine. Journal of Chemical Physics 21:1087–1091

MEYER, R., AND R. B. MILLAR. 1999. Bayesian stock assessment using a state-space implementation of the delay difference model. Canadian Journal of Fisheries and Aquatic Sciences 56: 37–52.

NICHOLS, J. D., G. L. HENSLER, AND P. W. SYKES, JR. 1980. Demography of the Everglade kite: implications for population management. Ecological Modelling 9:215–232.

———, J. E. HINES, J.-D. LEBRETON, AND R. PRADEL. 2000. Estimation of contribution to population growth: a reverse-time capture–recapture approach. Ecology 81:3362–3376.

ROSE, K. E., T. H. CLUTTON-BROCK, AND F. E. GUINNESS. 1998. Cohort variation in male survival and lifetime breeding success in red deer. Journal of Animal Ecology 67:979–986.

SARGENT, D. J. 1998. A general framework for random effects survival analysis in the Cox proportional hazards setting. Biometrics 54:1486–1497.

SCHLUTER, D., AND L. GUSTAFSSON. 1993. Maternal inheritance of condition and clutch size in the collared flycatcher. Evolution 47:658–667.

SEDINGER, J. S., P. L. FLINT, AND M. S. LINDBERG. 1995. Environmental influence on life-history traits: growth, survival, and fecundity in black brant (*Branta bernicla*). Ecology 76:2404–2414.

SERVICE, P. M. 2000. Heterogeneity in individual mortality risk and its importance for evolutionary studies of senescence. American Naturalist 156:1–13.

SPIEGELHALTER, D. J., A. THOMAS, N. G. BEST, AND W. R. GILKS. 1995. BUGS: Bayesian inference using Gibbs sampling. Version 0.50. MRC Biostatistics Unit, Cambridge, United Kingdom.

STEARNS, S. C. 1992. The evolution of life histories. Oxford University Press, New York, USA.

VAUPEL, J. W., AND A. I. YASHIN. 1985*a*. Heterogeneity's ruses: some surprising effects of selection on population dynamics. American Statistician 39:176–185.

———, AND ———. 1985*b*. The deviant dynamics of death in heterogeneous populations. Pages 180–211 *in* N. Brandon Tuma, editor. Sociological methodology. Jossey-Bass, San Francisco, California, USA.

VOUNATSOU, P., AND A. F. M. SMITH. 1995. Bayesian analysis of ring-recovery data via Markov chain Monte Carlo simulation. Biometrics 51:687–708.

WHITE, G. C. 2000. Population viability analysis: data requirements and essential analyses. Pages 288–331 *in* L. Boitani and T. K. Fuller, editors. Research techniques in animal ecology. Columbia University Press, New York, USA.

Received 30 March 2000.
Accepted 24 October 2001.
Associate Editor: Udevitz.