Ecological Applications, 30(5), 2020, e02112

A need for speed in Bayesian population models: a practical guide to marginalizing and recovering discrete latent states

Charles B. Yackulic,[1,4] Michael Dodrill,[1] Maria Dzul,[1] Jamie S. Sanderlin,[2] and Janice A. Reid[3]

[1]*Southwest Biological Science Center, U.S. Geological Survey, 2255 North Gemini Drive, Flagstaff, Arizona 86001 USA*
[2]*USDA Forest Service, Rocky Mountain Research Station, Flagstaff, Arizona 86001 USA*
[3]*USDA Forest Service, Pacific Northwest Research Station, Roseburg Field Station, Roseburg, Oregon 97331 USA*

Citation: Yackulic, C. B., M. Dodrill, M. Dzul, J. S. Sanderlin, and J. A. Reid. 2020. A need for speed in Bayesian population models: a practical guide to marginalizing and recovering discrete latent states. Ecological Applications 30(5):e02112. 10.1002/eap.2112

Abstract. Bayesian population models can be exceedingly slow due, in part, to the choice to simulate discrete latent states. Here, we discuss an alternative approach to discrete latent states, marginalization, that forms the basis of maximum likelihood population models and is much faster. Our manuscript has two goals: (1) to introduce readers unfamiliar with marginalization to the concept and provide worked examples and (2) to address topics associated with marginalization that have not been previously synthesized and are relevant to both Bayesian and maximum likelihood models. We begin by explaining marginalization using a Cormack-Jolly-Seber model. Next, we apply marginalization to multistate capture–recapture, community occupancy, and integrated population models and briefly discuss random effects, priors, and pseudo-R^2. Then, we focus on recovery of discrete latent states, defining different types of conditional probabilities and showing how quantities such as population abundance or species richness can be estimated in marginalized code. Last, we show that occupancy and site-abundance models with auto-covariates can be fit with marginalized code with minimal impact on parameter estimates. Marginalized code was anywhere from five to >1,000 times faster than discrete code and differences in inferences were minimal. Discrete latent states and fully conditional approaches provide the best estimates of conditional probabilities for a given site or individual. However, estimates for parameters and derived quantities such as species richness and abundance are minimally affected by marginalization. In the case of abundance, marginalized code is both quicker and has lower bias than an *N*-augmentation approach. Understanding how marginalization works shrinks the divide between Bayesian and maximum likelihood approaches to population models. Some models that have only been presented in a Bayesian framework can easily be fit in maximum likelihood. On the other hand, factors such as informative priors, random effects, or pseudo-R^2 values may motivate a Bayesian approach in some applications. An understanding of marginalization allows users to minimize the speed that is sacrificed when switching from a maximum likelihood approach. Widespread application of marginalization in Bayesian population models will facilitate more thorough simulation studies, comparisons of alternative model structures, and faster learning.

Key words: augmentation; autologistic; closed conditional; density dependence; forward conditional; fully conditional; hidden Markov model; mark–recapture; N*-occupancy; unconditional.*

Introduction

Bayesian implementations of population models are becoming increasingly common but can be exceedingly slow when confronted with large data sets. Use of Bayesian population models in open software has been facilitated by a series of recent books presenting intuitive code that allows users to modify and extend models (Royle and Dorazio 2008, Kéry 2010, Kéry and Schaub 2011). While flexible, the most popular approaches to implementing Bayesian population models are often orders of magnitude slower than maximum likelihood approaches. Slow models may be acceptable when they take minutes or a few hours, but speed has practical consequences if models take days to weeks to converge, a feature that is not uncommon when data sets are large and/or sparse. As scientists working with management agencies, we have, at times, adopted maximum likelihood approaches when we would have preferred a Bayesian approach simply because models were not on track to converge in a timeframe that would be relevant to management decisions. We have also reviewed articles for journals where authors suggested that the slowness of their model precluded simulation studies associated with new models (potentially leading to the adoption of methods that have not been fully tested), or led them to

Manuscript received 25 October 2019; accepted 24 January 2020; final version received 20 February 2020. Corresponding Editor: Beth Gardner.
[4] E-mail: cyackulic@usgs.gov

only consider a limited set of potential models (e.g., dropping the number of covariates evaluated or not exploring different priors).

While common approaches to Bayesian population models can be slow, they also have several advantages over maximum likelihood approaches. These advantages include that it is straightforward to include random effects and prior information in Bayesian analyses, and it is easier to calculate derived quantities and their associated uncertainties. While statisticians can, and have, found approaches to incorporating some of these advantages in maximum likelihood frameworks, the process of deriving maximum likelihood solutions for a specific problem can itself be a slow process, even if the resulting computations are quick. An ideal approach for many applied statisticians and quantitative ecologists would combine the speed of maximum likelihood with the flexibility and generality of Bayesian approaches.

One factor that slows common Bayesian approaches is the approach that is taken to handle discrete latent states. The concept of discrete latent states is ubiquitous in the capture–recapture and occupancy literatures and has been commonly applied in maximum likelihood approaches for decades. Relatively simple models such as a Cormack-Jolly-Seber (hereafter CJS; Cormack 1964) open population model or a single-season occupancy model can be formulated as discrete latent state models in which the states are alive and dead (CJS model) or occupied and unoccupied (occupancy model). In both cases, capture or detection leads to a definitive state assignment (alive or occupied), whereas nondetection admits state uncertainty (it is unknown whether the individual is still alive or whether the site was occupied).

The greatest benefits of discrete latent states, however, have been in their general application to more complex ecological questions via multistate models (Arnason 1973, Brownie et al. 1993, Schwarz et al. 1993, Lebreton et al. 2009). In capture–recapture analyses, multistate models are used to make inferences about a variety of ecological parameters. For example, if discrete locations are defined as states, inferences can be made about spatial variation in vital rates and detection probability as well as movement between locations. Similarly, multistate occupancy can be used to answer a variety of questions including how competition alters the colonization and extinction probabilities of competing species. In models that deal with repeated counts, such as the *N*-mixture model, the discrete states are the number of individuals in a site and each site can theoretically be in any non-negative integer state (i.e., 0, 1, 2, …, ∞). In each of these instances, the discrete states are referred to as latent because they are imperfectly observed (i.e., partially hidden). Individuals may be present, but not captured, or can move to unsampled sites, one or both species may go undetected when both are present in an area, and the number of individuals counted at a site will often be less than are present. However, it is not discrete latent states per se that slows Bayesian population analyses, rather it is how they are implemented.

Here, we discuss marginalization, an approach to implementing discrete latent states that can be orders of magnitude quicker than the more commonly used approach in Bayesian population models. While marginalization is ubiquitous in maximum likelihood approaches and has been used in Bayesian contexts (Rankin et al. 2016, Itô 2017, Yackulic et al. 2018b), we are not aware of any detailed overview of this concept. Our manuscript has two goals: (1) to introduce readers unfamiliar with marginalization to the concept and provide worked examples and (2) to discuss topics of interest to those already familiar with marginalization that, to our knowledge, have not been formally discussed in detail. To achieve these goals, we divide the remainder of the manuscript into five sections. We begin with an introduction to the concept of marginalization, illustrating the minor differences in coding required to marginalize a simple CJS model, and showing the significant increases in speed that ensue. In the second section, we apply marginalization to additional classes of models and discuss factors that motivate us to take Bayesian approaches in some of our work. The third and fourth sections will be of more interest to those already familiar with marginalization. The third section focuses on recovery of discrete latent states, including different definitions of conditional probabilities and the calculation of derived quantities such as population abundance or species richness. The fourth section discusses the use of auto-covariates in marginalized code. Auto-covariates are used when the state of a system at a given time is expected to affect a vital rate in the subsequent interval and include formal tests of density dependence (i.e., abundance affects survival or recruitment) or autologistic effects (i.e., the number and spatial arrangement of occupied neighbors affect probabilities of site colonization and extinction). There is a common misconception that discrete latent states are required for models including auto-covariates to be unbiased. We finish with a short fifth section summarizing the main points of our manuscript, including a synthesis of the observed improvement in computational speed across applications.

Throughout the first four sections, we present applications to real data to illustrate key concepts and provide speed comparisons between discrete (i.e., simulated) and marginalized implementations in various open-source software. While we compare performance across these software programs, the general usefulness of marginalization applies even if users are writing their own Markov Chain Monte Carlo (MCMC) samplers or using programs not specifically tested here. All code is provided in Data S1, which include details on the versions of various programs and packages used in our analyses. Data required to run applications are all available from the USGS ScienceBase Catalog (Yackulic 2018, Yackulic et al., 2018a, 2020). All speed comparisons were done on

servers with the same technical specifications and are based on fitting the same model 10 times. All Bayesian models were run in parallel. Models differ in the number of iterations required to reach convergence and we chose the number of iterations that were expected to lead to at least 50 effective samples across all monitored parameters. Since programs differ in how they calculate effective samples internally, we exported posteriors from each model and calculated effective samples using the coda package (Plummer et al. 2006) for all comparisons. We quantified speed in terms of the time elapsed per 100 effective samples in the least converged parameter (hereafter referred to as speed). Initial analyses suggested this approach to quantifying speed was relatively robust to the number of effective samples, provided the model had at least 10 effective samples for each parameter and the minimum number of effective samples was less than the maximum stored by a particular program (Appendix S1: Fig. S1).

How to Marginalize Discrete Latent States

Under the assumptions of closure common to most population models, individuals or sites are in a single discrete state at a given time. In a CJS model, an individual is either dead or alive and, in an occupancy model, a site is either occupied or unoccupied. However, these underlying (latent) states are observed imperfectly leading to uncertainty. Discrete latent state approaches faithfully represent this condition by simulating the underlying state of a system and the detection process conditional on the underlying state. In a given iteration of a Bayesian sampler, an individual may be dead or alive at a given time or a site may be occupied or unoccupied at a given time. Marginalized approaches, in contrast, do not simulate the "true" underlying state, but rather track the likelihood of being in any given state. While marginalized approaches involve keeping track of more information in each iteration, they also efficiently cover all state possibilities and typically converge in fewer iterations. When individuals or sites share capture or detection histories, marginalized approaches also allow users to only calculate the likelihood a single time in each iteration, rather than requiring each individual or site to be represented separately.

To make the distinction between discrete latent states and marginalized latent states clearer, we next consider a simplified CJS example. Our imaginary data set, **Y**, consists of a matrix in which rows represent 10 marked individuals released at time $t = 0$, sampled at time $t = 1$, and sampled again at time $t = 2$ and columns represent whether individuals were observed in each time step (Table 1). To simplify later generalization to multistate problems, we use 1's to indicate that an individual was observed alive, and 2's to indicate that an individual was not observed. Using discrete latent states, we would simulate whether each individual was alive (1) or dead (2) during each period in each posterior draw, based on the survival probability, φ, yielding a matrix of discrete latent states, **Z**. Note that some values in **Z** are constrained to equal 1 (indicated by underlining in Table 1) because **Y** indicates the individual was alive during that sampling event, or in a following sampling event (i.e., the second detection history, 121, can only occur if an individual survives both intervals). Other values, however, can and would vary in each iteration of the Markov Chain Monte Carlo (MCMC) simulation. The likelihood of the data, **Y**, conditional on the discrete latent states, **Z,** and the capture probability, p, is then given by the multinomial distribution with a vector of probabilities equal to $[p\ 1 - p]$ when the corresponding value in **Z** is 1 (indicating the individual is alive) and [0 1] when the corresponding value in **Z** is 2 (indicating the individual is dead and can only be unobserved).

Table 1. Simulation approach to discrete latent states.

	Z: simulated discrete latent states (different MCMC iterations)			
Observed capture histories	#1	#2	#3	...
111	**111**	**111**	**111**	...
121	**111**	**111**	**111**	...
112	**11**2	**11**1	**11**1	...
112	**11**1	**11**2	**11**1	...
122	**1**11	**1**12	**1**12	...
122	**1**22	**1**12	**1**22	...
122	**1**12	**1**22	**1**22	...
122	**1**12	**1**11	**1**11	...
122	**1**22	**1**22	**1**11	...
122	**1**11	**1**12	**1**11	...

Notes: The 1's in capture histories indicate that an individual was observed alive, and 2's indicate that an individual was not observed. The 1's in simulated discrete latent state indicate an individual was alive and 2's indicate an individual was dead. Discrete latent states are simulated in each iteration of a Bayesian simulation algorithm. Underlined and boldface values are constrained by the observed capture history, but other values vary in each iteration. The likelihood of the observations is calculated conditional on survival and detection probabilities as well as the simulated states (**Z**). The ellipses (...) are meant to signify that there would be many more iterations run for a given MCMC chain. MCMC, Markov Chain Monte Carlo.

Using a marginalized approach, we would calculate the likelihood of a given capture history by keeping track of the likelihood of being in either state at each time interval conditional on the capture history and the parameters, p and φ, and iterating through each time step. To keep track of these likelihoods, we introduce an array, ζ, which has three dimensions (one defined by the number of unique capture histories, a second by the number of capture events, and a third dimension defined by the number of possible states). As mentioned above, this approach requires keeping track of more information, but more efficiently samples all the possible chains of states. The likelihood of each capture history can then be calculated by summing the entries from the final time step ($t = 3$ in Table 2), yielding a likelihood of $p^2\varphi^2$ for an individual that was seen during both recapture events

Table 2. Marginalization approach to discrete latent states.

Observed unique capture histories	Frequency	ζ: marginalized calculations at each time step	
		$t = 2$	$t = 3$
111	1	$\begin{bmatrix} p\varphi \\ 0 \end{bmatrix}$	$\begin{bmatrix} p^2\varphi^2 \\ 0 \end{bmatrix}$
121	1	$\begin{bmatrix} (1-p)\varphi \\ (1-\varphi) \end{bmatrix}$	$\begin{bmatrix} (1-p)p\varphi^2 \\ 0 \end{bmatrix}$
112	2	$\begin{bmatrix} p\varphi \\ 0 \end{bmatrix}$	$\begin{bmatrix} p(1-p)\varphi^2 \\ p\varphi(1-\varphi) \end{bmatrix}$
122	6	$\begin{bmatrix} (1-p)\varphi \\ (1-\varphi) \end{bmatrix}$	$\begin{bmatrix} (1-p)^2\varphi^2 \\ (1-\varphi)(1+(1-p)\varphi) \end{bmatrix}$

Notes: The 1's in capture histories indicate that an individual was observed alive, and 2's indicate that an individual was not observed. The probability of being in each state is calculated in each time step base on the observed capture history and survival (φ) and detection (p) probabilities. The overall likelihood of an observation is given by the sum across all states in the last time step and the log of this likelihood is multiplied by the frequency of a unique capture history and summed across all unique capture histories.

(the first capture history in **Y**), a likelihood of $(1 - p)$ $p\varphi^2$ for an individual seen only on the second recapture event (the second capture history in **Y**), $p\varphi\ (1 - \varphi) + p$ $(1 - p)\varphi^2$ for an individual seen only on the first recapture event (third unique capture histories in **Y**) and $(1 - \varphi) + (1 - p)\varphi(1 - \varphi) + (1 - p)^2\varphi^2$ for the remaining capture histories in which individuals were never recaptured. Importantly, because the discrete latent states are not actually simulated, it is not necessary to run identical capture histories multiple times, rather, the likelihood of a capture history can simply be raised to the power of its frequency (or frequency can be multiplied by the log-likelihood of a capture history).

*Application 1: CJS applied to rainbow trout (*Oncorhynchus mykiss*)*

For a more concrete illustration of these different approaches, we next apply a CJS model to real data. Data in this application come from a mark–recapture study of rainbow trout in the Colorado River just downriver of its confluence with the Little Colorado River. Rainbow trout were sampled using electrofishing on 18 quarterly trips between April 2012 and October 2016 (for more study details, see Korman et al. 2015, Yackulic et al. 2018b).

For this application only, we present the code in the manuscript text that distinguishes the two different approaches to illustrate the minor differences. We begin by fully describing code for discrete (D) and marginalized (M) versions of a fully time-specific CJS model in the BUGS language. We then give a brief overview of a M version in the Stan language (Carpenter et al. 2017). Stan is an increasingly popular (and faster) open software that does not support discrete latent states. We refer readers to the supplementary material for the full Stan code of this and other applications. Next, we fit D and M versions of the model in three popular open software programs that share the common BUGS language: JAGS (J; Plummer 2003), WinBUGS (W; Lunn et al. 2000), and OpenBUGS (O; Lunn et al. 2009). We compare the speed and output from these six options (JD, JM, WD, WM, OD, OM) to a seventh option consisting of a marginalized version fit in Stan (SM). We stress that marginalization can be applied in various programs and is not a specific feature of the programs we use here.

D and M code in the BUGS language share a first block of code in which survival, φ_t, and capture probability, p_t, are given uniform priors between zero and one. In addition, we introduce two arrays defining state transitions, ω, and the observation process, ρ. The first row in ω keeps track of the probability that an individual survives, φ_t, (i.e., remains in state 1) or dies, $1 - \varphi_t$, (i.e., transitions to state 2). The second row ensures that dead individuals remain dead (i.e., the transition from state 2 to 1 is fixed at 0 and the transition from state 2 to state 2 is fixed at 1). The dimensions of this array could easily be extended for a multistate mark–recapture model and values within the array modified for a dynamic occupancy model. The first row of ρ keeps track of the probability of being captured (p_t) or not captured ($1 - p_t$) conditional on being alive, whereas the second row ensures that dead individuals can only ever be not captured. Again, the dimensions could easily be extended for multistate models and values defined differently for different classes of models (e.g., false-positive models). The parameters ω and ρ are defined for each interval, t, in the set of total intervals, Nint

```
for (t in 1 : Nint){
   phi[t] ~ dunif(0,1)
   p[t] ~ dunif(0,1)
   omega[1, 1, t] = phi[t]
```

```
  omega[1,2,t] = (1 - phi[t])
  omega[2,1,t] = 0
  omega[2,2,t] = 1
  rho[1,1,t] = p[t]
  rho[1,2,t] = 1 - p[t]
  rho[2,1,t] = 0
  rho[2,2,t] = 1}.
```

In the D version, this block of code is followed by a second block in which **Y** and **Z** matrices are specified for each individual, *i*, and time, *t* according to

```
for (i in 1 : NY){
  Z[i, indf[i]] = 1
  for (t in indf[i] : Nint){
    Z[i, (t+1)] ~ dcat(omega[Z[i,t], , t])
    Y[i, (t+1)] ~ dcat(rho[Z[i, (t+1)], , t])}}
```

where NY is the number of individuals captured one or more times, **Z** is a matrix with the discrete latent states for each individual during each sampling event, **Y** is a matrix with capture histories for each individual, indf is a vector specifying the first trip on which each individual is captured, and dcat refers to the categorical distribution (analogous to the multinomial distribution). The code loops through all the individuals (i.e., from 1 to NY). Within each individual, it fixes the latent state, **Z**, to be alive (i.e., **Z** = 1) in the first time period they were observed, and then loops forward through the remaining time periods, simulating both the process of survival and death (through the **Z** matrix), and the detection process (by comparing the **Z** matrix to the actual capture history data found in the **Y** matrix).

In the M version, the first block of code remains the same, but the second block becomes

```
for (i in 1 : NS){
  zeta[i, sumf[i], 1] = 1
  zeta[i, sumf[i], 2] = 0
  for (t in sumf[i] : Nint){
    zeta[i, (t+1), 1] = inprod(zeta[i, t, ],
    omega[, 1, t]) * rho[1, y_sum[i, (t+1)], t]
    zeta[i, (t+1), 2] = inprod(zeta[i, t, ],
    omega[, 2, t]) * rho[2, y_sum[i, (t+1)], t]}
    lik[i] = sum(zeta[i, (Nint+1), ])
    fr[i] ~ dbin(lik[i], FR[i])}
```

where NS is the number of summarized (unique) capture histories, y_sum is a matrix with each unique capture history in Y, FR is the frequency of each capture history, sumf is a vector specifying the first trip with a capture history for each unique capture history, zeta is a three-dimensional array, lik is a vector with the likelihood of each unique capture history, dbin refers to the binomial distribution, inprod refers to the inner product of two vectors, and fr has the same values as FR. Here the code only cycles through unique capture histories. For each unique capture history, the code fixes the likelihood of being in the alive state to 1 and the likelihood of being in the dead state to zero. The code then loops forward through time using matrix multiplication to determine the likelihood of being in each state at time $t + 1$, given information encapsulated in the ζ array and the transition matrix ω, and standard multiplication with the probability of the observation in time $t + 1$. An important feature of this code is that if an individual is observed in time period $t + 1$ (i.e., y_sum[i, $t + 1$] equals one), then $\rho_{2,1}$, the probability of observing a dead individual as alive is zero and thus the overall probability of being in the dead state in that interval becomes zero. Once the code has looped through all time periods, the likelihood of that individual capture history is then calculated by summing across all possible states. The final line of the code is a modification of the "ones" trick in the BUGS language, which updates the posterior by the value of the likelihood raised to its frequency.

The Stan language differs from BUGS language in several ways that have been reviewed elsewhere (Betancourt 2017, Monnahan et al. 2017) and we provide code in the supplementary materials, so readers can familiarize themselves. However, code for a SM version of the CJS model differs from the M version presented above in only one substantial way. Namely, the last two lines of the M version in the BUGS language are replaced by the following line of code:

```
target += FR[i] * log(sum(zeta[i, (Nint + 1),]))
```

which directly adds the log-likelihood of a unique capture history to the posterior.

Different versions of the same model led to indistinguishable estimates of parameter means and quantiles but varied by almost four orders of magnitude in the time required to produce 100 effective samples of the least converged parameter. Discrete versions in JAGS, WinBUGS, and OpenBUGS took a few hours, while marginalized versions in the same languages took a few minutes and a marginalized version in Stan took a few seconds (Fig. 1). Differences in speed between versions were attributable to two factors: (1) the degree of autocorrelation in posterior draws, which determined how many iterations were required per effective sample and (2) the average speed of each iteration. Discrete versions generally led to highly correlated posterior draws requiring ~200 iterations per effective sample, whereas marginalized versions in the BUGS language required ~10 iterations per effective sample and the SM version required fewer than 2 iterations per effective sample (Appendix S1: Fig. S2). SM was also the quickest version at producing iterations yielding ~60 posterior samples per second, whereas the JM yielded ~12 samples per second, and other versions yield 1.5–4.0 samples per second. We fit all seven versions over a range of iterations and found our conclusions were insensitive to the total number of iterations that models were run provided models were somewhat converged (Appendix S1: Fig. S1). For the rest of the manuscript, we focus on JD and JM versions of model as proxies for the other BUGS language versions alongside SM versions. To provide

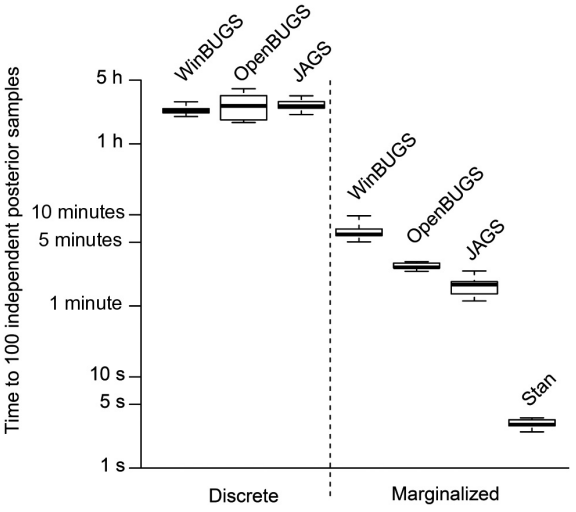

FIG. 1. Comparison of the speed of seven different versions of a Cormack-Jolly-Seber (CJS) model applied to rainbow trout data. Whiskers cover the full range of observed speeds based on 10 runs of each version of the CJS model. Speed is measured as the time per 100 effective samples in the least converged parameter.

further illustration of the need for faster Bayesian population models, we next consider a more complex model.

*Application 2: Multistate mark–recapture model applied to endangered humpback chub (*Gila cypha*) in the Colorado River (CR) and its tributary the Little Colorado River (LCR)*

This application was the primary motivation for many of the authors to first explore marginalization in a Bayesian framework. We have yet to run a JD version of the model for enough time to reach convergence. The data are sparse and include large numbers of animals that are marked and never seen again. Data were collected between 2009 and 2017 in both the LCR and the fixed site in the CR. In addition to three spatial states (LCR, CR in the sampled area, CR outside the sampled areas), we considered two size classes referred to as small adults (200–249 mm) and large adults (250+ mm) because of differences in the vital rates and detection probabilities of different size adults, leading to a total of six potential "alive" states for adults in the multistate mark–recapture model. We ran our model on a seasonal basis. Biological parameters include size transition (growth) and location transition (movement) parameters, as well as survival and a parameter that represents the proportion of adult humpback chub found in the CR study area out of the total population of adult humpback chub that live in the CR and spawn in the LCR. The sampled and unsampled portions of the CR are assumed to have the same survival, growth, and movement parameters, however, survival, growth, and movement parameters do differ between size classes and between the LCR and CR. Survival is assumed to be constant with respect to time, and growth and movement vary seasonally, but are constant across years. For more details, we refer readers to the supplementary materials and previous analyses reported in Yackulic et al. (2014b).

We attempted to fit these data using JD, JM, and SM versions of the code; however, the JD version was run for 20,000 iterations, taking 8.3 d (~700,000 s) and was still far from convergence (effective sample of 5.6 for least converged parameter). In contrast, all 10 runs of the SM version finished 2,000 iterations in less than two hours, yielding more than 500 effective samples for all parameters and an average time of ~15 minutes per 100 effective samples of the least converged parameter. JM was an order of magnitude slower at producing 100 effective samples of the least converged parameter but was still capable of yielding inferences on the scale of hours. While this section illustrates the motivation for marginalizing discrete latent states in a Bayesian framework, pragmatists may question why we would not just use a maximum likelihood framework, which would be even quicker. In fact, Yackulic et al. (2014b) used a maximum likelihood approach for this very reason; however, the assumption of temporally constant parameters is problematic, and the data are too sparse to fit the fully time-dependent model, thus motivating the use of random effects. In the next section, we return to this application illustrating how random effects can easily be incorporated into this model to provide inferences that are not easy to derive using a classic maximum likelihood approach.

To Bayes or Not to Bayes: A Pragmatic Response

In our experience, most ecologists (not to mention stakeholders and resource managers) are not making decisions about whether a Bayesian or maximum likelihood approach is appropriate based on philosophical arguments. Rather, choices frequently reflect a variety of factors, including familiarity with different approaches, features of a given problem, and perceived tradeoffs between the speed of a more complex analysis and the differences in value of inferences to be derived from a more complex analysis. It is not uncommon to read manuscripts purporting that a more complex analysis is better because it more closely mimics the complexity of ecological systems without clearly articulating actual advantages in terms of inferences. A pragmatic analyst does not choose a slower or more complex analysis if it will give identical, or near identical inferences (although they might reasonably choose an analysis that is slower computationally but does not require investing resources in developing new code).

From the perspective of a pragmatic analyst that is familiar with maximum likelihood, there is no reason to take a Bayesian approach for either of the first two applications. The priors on all parameters are uniform

and cover the full range of sensible values (i.e., 0 to 1) so a maximum likelihood approach would provide indistinguishable parameter estimates. Furthermore, maximum likelihood estimation is almost always quicker than even a marginalized Bayesian approach, and with minimal modifications, the code presented for these applications can be used to fit via maximum likelihood. However, there are benefits to taking a Bayesian approach for some applications. In the rest of this section, we explore three applications where there is value in taking a Bayesian approach over a maximum likelihood approach. Specifically, we focus on the added inferential value to be gained from random effects and informative priors.

Revisiting application 2 to make inferences about temporal variation in vital rates

In application two, adult humpback chub survival was assumed to be constant and movement and growth were assumed to differ seasonally, but to be constant across years. Some authors would refer to this as a model with full pooling of information (or nearly full because of the seasonal differences). Attempts to relax these assumptions and estimate different rates for each time interval (a time-dependent parameterization like application 1) would be referred to as a complete lack of pooling. For these data, a time-dependent parameterization leads to uselessly imprecise estimates with 95% confidence intervals spanning essentially from 0 to 1. Random effects represent an intermediate between these two extremes sometimes referred to as partial pooling (Gelman and Hill 2007). With rich data sets (or when the number of intervals are less than five) random effects models will behave similarly to time-dependent models. With sparser data and more intervals, however, random effects models will behave more like a constant parameter model. Benefits of random effects in population models have been recognized for almost two decades (Link 1999, Burnham and White 2002, Royle and Link 2002). While random effects can be included in maximum likelihood analysis, their application is often not straightforward, especially when random effects are included in more than one parameter of interest. One advantage of a partial pooling approach is that in some situations, it can allow us to detect temporal changes in vital rates that would be obscured by imprecision in a time-dependent model and assumed to be nonexistent in a fully pooled model.

In the case of humpback chub, a substantial decline in adult catch in the LCR spatial state was observed beginning in spring 2015 and there were questions about the degree to which this decline reflected variation in survival vs. movement rates with implications for the status and trend of the overall population. To quantify variation in these parameters, we fit a version of the marginalized model in Stan with random effects for survival, growth and movement (hereafter referred to as the SMRE version of the multistate model). Addition of random effects increased the time to sample 100 effective samples by 31% primarily by modestly increasing autocorrelation in posterior draws. Addition of random effects illustrated that the declines in LCR catch were the result of reduced adult movement into the LCR during the spawning season in 2015 as opposed to declines in adult survival (Fig. 2A, B). Consistent with this demographic interpretation, humpback chub catch in the LCR has recovered in recent years and estimates of total adult abundance were stable throughout the study period (Fig. 2C).

In many instances, it is advisable to consider time-varying fixed effects based on well-reasoned a priori hypotheses alongside temporal random effects. As an example, we included season as a factor in both the growth and movement portions of the humpback chub application based on a priori knowledge that these rates vary substantially by season. When data are sparse, and the uncertainty associated with any given interval is high, there is potential for an important aberration in a vital rate over one or more intervals to be obscured by pooling if fixed effects are not also included. Using random effects alongside fixed time-varying covariates is also advisable as an approach to deal with the nonindependence of individuals exposed to the same conditions during a given interval (Barry et al. 2003).

Pseudo-R^2

Use of random effects, alongside fixed effects, also allows ecologists to quantify how much of the variation in one or many parameters of interest is explained by a covariate or series of covariates, by calculating a pseudo-R^2 value. This approach differs from an analysis of deviance, which is applied to nested models and focuses on the overall fit of a model. Calculation of pseudo-R^2 values follows from Gelman and Pardoe's (2006) treatment of multi-level R^2. Specifically, the pseudo-R^2 for a parameter is defined as

$$\text{pseudo} - R^2 = 1 - \frac{E(\text{var}_{\text{RE}})}{E(\text{var}_{\text{RE+FE}})}$$

where $E(\text{var}_{\text{RE}})$ is the expected variance of random effects across posterior draws and $E(\text{var}_{\text{RE+FE}})$ is the expected variance of the sum of the random effects and all fixed effects across posterior draws. While this approach is intuitive, it has not been widely used in the population modeling literature (but see Yackulic et al. 2018b). In the humpback chub application, season is included as a factor in the equations governing size transitions (i.e., growth) in both spatial locations, and there are four separate movement parameters estimated in each interval so we can estimate how much of the temporal variance in each of these parameters of interest is explained by season. We calculate pseudo-R^2s of 0.79 and 0.92 for movement of small and large adult

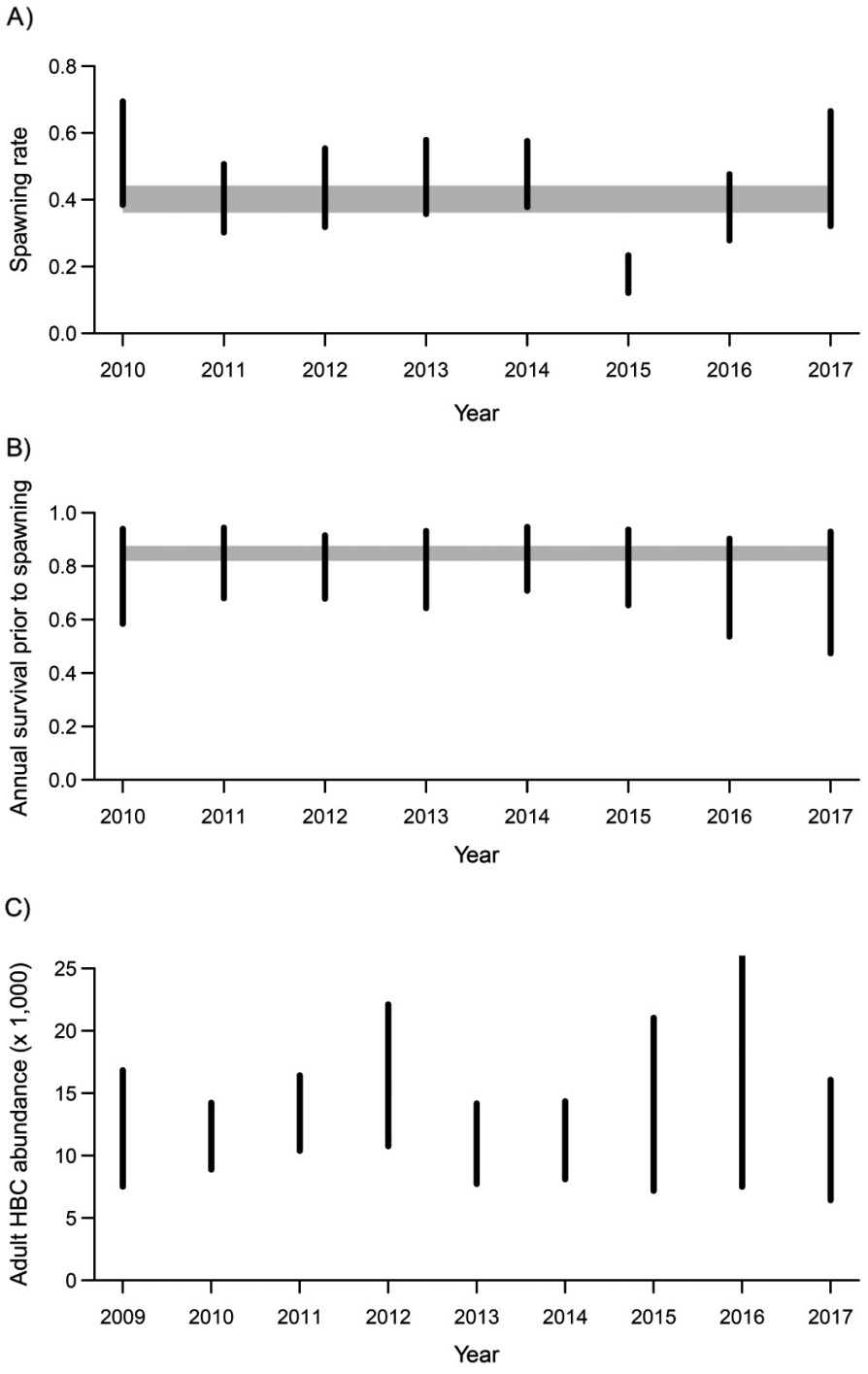

FIG. 2. Estimates of (A) movement into the Little Colorado River by large adult humpback chub during the spawning season (spawning rate), (B) annual survival of large adults in the Colorado River, under a model with random effects and a model that assumes constant rates, and (C) total abundance of adult humpback chub that spawn in the Little Colorado River. Lines indicate the 95% credible intervals for estimates from each year given by the random effect model, while the gray polygon represents the 95% credible intervals from a model that assumes a constant rate over time.

humpback chub out of the CR into the LCR, and 0.66 and 0.78 for movement of small and large adult humpback chub out of the LCR into the CR. With respect to growth, season explains 0.77 of the variance in the CR and 0.91 of the variance in the LCR. In the next application, we consider a situation where random effects are used to pool among species in a community as opposed to over temporal intervals.

Application 3: Dynamic community occupancy model applied to birds in the Sky Islands of Arizona

Community occupancy models are an extension of single-species occupancy models (Dorazio et al. 2006) and are commonly applied to data from avian studies that use point-count sampling (Russell et al. 2009). Community occupancy models estimate occupancy, species richness, and with multiple time periods, local colonization and extinction probabilities, while accounting for imperfect detection. Despite the growing popularity of community occupancy models, the slowness of these models has led many applications to explore only a limited set of potential competing models, and has limited the scope of simulation studies exploring issues of study design and biases (but see Sanderlin et al. 2014, Guillera-Arroita et al. 2019). Here we illustrate how community occupancy models can be marginalized using a model with multiple years of data in which species' detection probabilities are modeled as random effects and species' colonization and extinction probabilities are modeled using random effects that also vary by seven habitat types. The complexity of this model is representative of commonly applied community occupancy models.

We apply the model using detection/nondetection data collected for 149 bird species across 92 sites in the Chiricahua Mountains of southeastern Arizona, USA. Each site was visited three times in each of five years (1991–1995), with the exception that only 68 sites were visited in the first year of the study. These data are a subset of data previously analyzed by Sanderlin et al. (2013). To make the example more tractable (but still cover the range of bird species that are abundant and easy to detect vs. rare and hard to detect), we selected the largest mountain range in this study, the Chiricahua Mountains, and only made inference on species detected at least once in the original study (i.e., we included species not detected in the Chiricahua Mountains, but detected in another mountain range). For more details concerning the model and data, see Sanderlin et al. (2013) and the supplementary materials.

JD, JM, and SM versions lead to indistinguishable estimates of all parameters we monitored. Interestingly, the JM version was the fastest taking an average of 17 minutes to reach a minimum of 100 effective samples, while the SM version took an average of approximately 37 minutes and the JD averaged over 5 h. When models are applied to larger communities, we expect that the relative difference between M and D versions will likely increase, particularly, when data include many species that are only seen rarely and include many sites. Later in the manuscript we report estimates of species richness produced by marginalized and discrete versions of this model, however, we next consider an example that illustrates how informative priors can be useful when data are particularly sparse.

*Application 4: Informative priors in an integrated population model of brown trout (*Salmo trutta*)*

Ecologists do not frequently use informative priors intentionally (but see Garrard et al. 2012). Most Bayesian population analyses instead choose priors that are believed to be weakly informative or entirely noninformative, leading to estimates that are indistinguishable from a maximum likelihood approach. This approach is problematic in some situations for at least two reasons. First, seemingly uninformative priors can still affect inferences (Seaman et al. 2012, Northrup and Gerber 2018) and there is a tendency to not critically evaluate priors when they are believed to be uninformative (especially if models are very slow). Second, in applied settings, there are often situations where data are sparse, outside knowledge exists to inform one or more parameters, and the purpose of the modeling is not to test this existing knowledge, but to make predictions or test hypotheses about other parameters. Under these circumstances, failure to include prior information can lead to imprecise or biased predictions and inefficient management decisions. In this application, we show how priors and marginalization can lead to better, and faster, predictions to inform a decision-making process motivated by an increasing invasive species.

Brown trout have been present in the CR system for decades; however, catch of brown trout in the river segment just downstream of Glen Canyon Dam was low until a substantial increase between 2012 and 2017. Stakeholders and resource managers were concerned with this trend because brown trout are piscivorous and capable of large movements and a large brown trout population potentially threatens the population of humpback chub located downriver. The model presented here was developed to forecast the potential consequences of different management actions through a structured process described in Runge et al. (2018). The model assumes three size classes, a seasonal time step, and includes both fixed and random effects. The model is fit using catch and effort data collected between fall 2000 and fall 2017 throughout the ~25 km of the Glen Canyon reach of the CR, as well as mark–recapture data collected on seasonal trips between spring 2012 and fall 2017 at a fixed 2-km section within the larger reach. The mark–recapture data are sparse, and we realized during model development that we would not be able to provide useful forecasts without utilizing outside information in the form of priors. In particular, we used the Lorenzen equation (Lorenzen 2000) and estimates of the average length in each size class during each season, to derive a prior of seasonal and size class-specific survival rates. Past mark–recapture studies of other fish species in the CR have found substantial temporal variation in capture probability (Korman and Yard 2017, Yackulic et al. 2018b), therefore the model uses random effects to allow for temporal variation in

capture probability. Here, we fit four versions of the model, including JD, JM, and SM versions of the model with informative priors to compare computation speed. To illustrate the importance of informative priors in this application, we also compare the parameter estimates from the SM version to an SM version that replaces the informative prior on survival with an uninformative prior and removes random effects on capture probability (SM-NPI—No Prior Information).

Parameter estimates derived from SM, JM, and JD versions were indistinguishable. The time to reach a minimum 100 effective samples averaged 29, 161, and 193 minutes for SM, JM, and JD versions, respectively. Many estimates from SM and SM-NPI versions differed substantially. For example, the SM-NPI estimated annual survival of brown trout in the largest size class as 0.37 (95% CI: 0.2–0.54), which is substantially lower than estimates of rainbow trout survival in the same reach based on more robust data (Korman et al. 2015) as well as the estimate of 0.65 (95% CI: 0.41–0.85) for brown trout in the middle size class. At the same time, estimates of capture probability from a model without informative priors on survival were greater than for similar sized rainbow trout, even though adult brown trout are known, from unpublished telemetry studies in the system, to spend much of their time at depths that should make the relative inaccessible to electrofishing (R. Schelly, *personal communication*). Use of informative priors on survival and addition of random effects to capture probability led to estimates that were more consistent with these observations and also led to very different inferences regarding brown trout population dynamics (Fig. 3). While it would certainly have been preferable to have more robust data to base inferences on, we were inclined to put more weight on abundance estimates from the SM version since parameter estimates from the SM version made more ecological sense based on our understanding of how survival changes with body size in fish species in general, and based on expectations for capture probability in a large river such as the Colorado River.

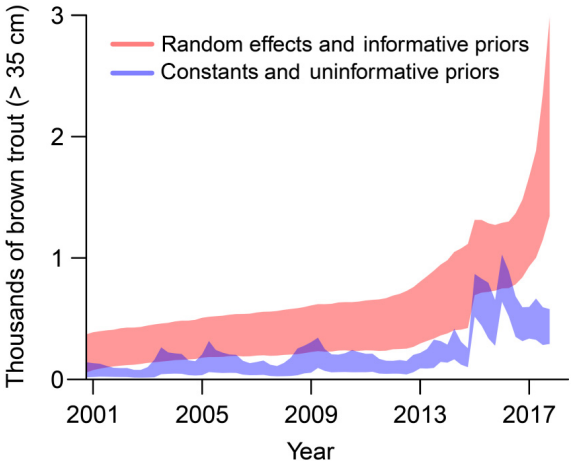

FIG. 3. Estimates of the abundance of brown trout over 35 cm over time in the Glen Canyon reach of the Colorado River under a model that includes random effects and informative priors or a model that assumes constant rates and uninformative priors. Constant models without informative priors suggest a volatile population, but also lead to estimates of other parameters that are unlikely. Inclusion of random effects and informative priors lead to parameter estimates closer to expectations and higher abundance estimates.

Summary

Understanding how to marginalize is useful for applying maximum likelihood approaches and improving the speed of Bayesian approaches. We generally choose an approach based on the features of a given problem. In this section, we have examined some applications for which a Bayesian approach is beneficial. In all three applications explored in this section, marginalization increased computational speed dramatically, which was important given that use of random effects generally slows models. For application 2, we previously showed that a JD version without random effects failed to converge after 8.3 d, so a random effects version was only possible through marginalization. In the next two sections of the manuscript, we discuss issues that are common to both maximum likelihood and Bayesian application of marginalization: the recovery of latent states and use of auto-covariates.

Recovering Discrete Latent States: Issues and Terms

One perceived issue with marginalizing discrete latent states is that inferences are sometimes focused on the discrete latent states. For example, it may be necessary to calculate the probability that a given site is occupied by a species or estimate the overall abundance of a population. Before explaining how latent states like these can be recovered, we begin by defining some general terms. We introduce the term *first-order state* to refer to a state at the scale of the individual or site (e.g., alive/dead, or occupied/unoccupied) and the term *second-order state* to refer to a state derived by applying some operation to the first-order states (e.g., abundance is the sum of alive individuals, species richness is the sum of all species occurring in a site). This distinction is important because if inferences are focused on second-order states only, it may not be necessary to recover first-order states, or it may be sufficient to imperfectly recover first-order states.

To clarify what is meant by imperfect recovery, we must first distinguish between unconditional and conditional estimates of first-order states and further distinguish between three different forms of conditional estimates. *Unconditional* estimates do not use specific information from an individual's capture history or site's detection history, relying instead on estimates of parameters (which are themselves informed by all the data) and individual or site covariates. A *closed* (or *seasonal*) *conditional* estimate combines the unconditional estimate with specific capture/detection information for

the period of interest; however, it ignores specific capture/detection information from preceding or subsequent time periods. A *forward conditional* estimate uses capture/detection information and relevant parameters from the current and all preceding time periods. Last, a *fully conditional* estimate uses the entire capture or detection history and relevant parameters to determine first-order state probabilities during any period. During the final period of a capture or detection history forward and fully conditional estimates are exactly the same. A fully conditional estimate should exactly match estimates derived using discrete latent states (a point we illustrate in the next paragraph). All other approaches should be imperfect estimators (imperfect in the sense that they do not use all the available information). However, these other approaches are easier to calculate and may be sufficient in many applications.

To make the distinction between these estimates of first-order states more concrete, consider an imaginary detection history from an occupancy study over three-time periods with two visits in each period [01 00 11]. We assume that initial occupancy ($\widehat{\psi}$), colonization ($\widehat{\gamma}$), extinction ($\widehat{\varepsilon}$), and detection probability ($\hat{p}$) are constant and previously estimated. We focus here on different estimates for the second period. The unconditional estimate for the second period would simply be the probability the site had been occupied in the first period, $\widehat{\psi}$, and did not go extinct, $1-\widehat{\varepsilon}$, during the first interval added to the probability the site was not occupied in the first period, $1-\widehat{\psi}$, and colonized, $\widehat{\gamma}$, during the first interval

$$\widehat{\psi}(1-\widehat{\varepsilon})+\left(1-\widehat{\psi}\right)\widehat{\gamma}\,(\text{unconditional})$$

Note that this estimate includes none of the specific detection information from any interval (including the information that we know the site was in fact occupied during the first interval) and only considers the probability the site was occupied. As soon as we begin to incorporate detection history information, we also need to consider the probability the site was not occupied. If we only use information from the second period, the site was either occupied and was not detected on both surveys with a probability of $[\widehat{\psi}1-\widehat{\varepsilon})+(1-\widehat{\psi})\widehat{\gamma}](1-\widehat{p})^2$ or was never occupied with a probability of $(1-[\widehat{\psi}(1-\widehat{\varepsilon})+(1-\widehat{\psi})\widehat{\gamma}])$. We can then calculate the closed conditional estimate by dividing the probability of occupancy by the sum of the probabilities of occupancy and nonoccupancy

$$\frac{[\widehat{\psi}(1-\widehat{\varepsilon})+\left(1-\widehat{\psi}\right)\widehat{\gamma}](1-\widehat{p})^2}{\left[\widehat{\psi}(1-\widehat{\varepsilon})+\left(1-\widehat{\psi}\right)\widehat{\gamma}\right](1-\widehat{p})^2+(1-\left[\widehat{\psi}(1-\widehat{\varepsilon})+\left(1-\widehat{\psi}\right)\widehat{\gamma}\right])}$$

(closedconditional)

Since the site was detected as occupied in at least one survey during the first period, we know the site was in fact occupied during the first period and can replace all $\widehat{\psi}$ in the closed conditional equation with 1, which leads to a much simpler expression for the forward conditional estimate

$$\frac{(1-\widehat{\varepsilon})(1-\widehat{p})^2}{(1-\widehat{\varepsilon})(1-\widehat{p})^2+\widehat{\varepsilon}}(\text{forwardconditional})$$

In other situations, we might also include a $(1-p)p$ term to reflect the detection history during the first period; however, this would end up in all terms in the numerator and denominator here and thus would cancel out. Similarly, when we consider the information from the third period it is tempting to include a p^2 in all terms to reflect the two detections; however, this would end up cancelling out. Instead the only relevant information from the full capture history is that the site must have not gone extinct, $(1-\widehat{\varepsilon})$, between the second and third periods if it was occupied in the second period, or must have been colonized, $\widehat{\gamma}$, if it was unoccupied during the second period. Taken together this yields the fully conditional estimate for the example

$$\frac{(1-\widehat{\varepsilon})(1-\widehat{p})^2(1-\widehat{\varepsilon})}{(1-\widehat{\varepsilon})(1-\widehat{p})^2(1-\widehat{\varepsilon})+\widehat{\varepsilon}\widehat{\gamma}}\,(\text{fullyconditional})$$

Taking the thought experiment a step further, we can calculate these different estimators under different parameter values and compare them to estimates from a discrete latent state approach (Table 3). Whereas the fully conditional estimate is always equivalent to the discrete latent state estimate, the other estimators are different and sometimes very different. For the parameter values considered here the forward conditional approach provides better approximations than the closed conditional or unconditional approaches.

More generally, unconditional, closed conditional and forward conditional estimates can be calculated easily with minimal modifications to the marginalized code presented for each application in this manuscript. For example, a forward conditional estimate of being in state, s, for any individual/site, i, and time, t, can be

TABLE 3. Estimates of occupancy during the second season of the three season detection history with two visits per season, [01 00 11], vary depending on parameter values and the approach to estimating occupancy.

ψ	γ	ε	p	U	closedC	forC	fullC	DLS
0.75	0.6	0.2	0.1	0.75	0.71	0.76	0.81	0.81
0.75	0.6	0.8	0.1	0.3	0.26	0.17	0.06	0.06
0.75	0.1	0.2	0.1	0.63	0.57	0.76	0.96	0.96
0.75	0.6	0.2	0.6	0.75	0.32	0.39	0.46	0.46

Notes: We assumed different values for model parameters: initial occupancy (ψ), colonization (γ), extinction (ε), and detection probability (p), and calculated unconditional (U), closed conditional (closedC), forward conditional (forC), and fully conditional (fullC) estimates and compared them to estimates from a discrete latent state model (DLS). All estimates are rounded to two decimal points.

calculated from the array, ζ as $\frac{\zeta_{i,t,s}}{\sum_{j=1}^{j=S}\zeta_{i,t,j}}$ where S is the total number of states. Deriving fully conditional estimates is less straightforward to calculate automatically as it involves calculating the likelihood of a particular capture or detection history for every possible sequence of discrete latent states. For this reason, it is useful to identify conditions when more easily calculated, but imperfect, estimates can be used instead of fully conditional estimates. In our experience, imperfect estimates of first-order states are sufficient when inferences are focused on second-order states. A trivial example illustrating this point is estimation of average occupancy across a population of sites from a single-season occupancy model with constant detection probability, p, and constant occupancy, ψ. Unconditional and conditional estimates (first-order states) will necessarily differ for each individual site, yet their sum (a second-order state) will be exactly equal.

For an example where estimates of first-order states can be skipped altogether, consider estimation of abundance from a closed capture–recapture model. In the Bayesian literature, abundance is frequently derived by simulating first-order latent states for individuals that were not encountered in a given sampling event via data augmentation (sometimes referred to as N-augmentation; Tanner and Wong 1987, Royle and Dorazio 2012) and then summing these first-order latent states to derive abundance, a second-order latent state. Alternative approaches to the data augmentation approach include reversible jump Markov chain Monte Carlo (RJMCMC; Green 1995, King and Brooks 2001, Schofield and Barker 2008) sampling and Monte Carlo integration within Markov Chain Monte Carlo (MCMC; see Bonner and Schofield 2014 for a comparison of these approaches). King et al. (2016) introduce a computationally more tractable approach using a semi-complete data likelihood approach, which can be used when there is individual heterogeneity in capture probability and alongside a variety of priors on abundance. A simple alternative that falls within their framework exists when capture probability is constant yet has not been widely applied (but see Yackulic et al. 2018b). This alternative involves estimating the number of uncaptured individuals in each time period using draws from the negative binomial distribution with the number of captured individuals in a particular time period serving as the number of successes, and the posterior draw from each MCMC iteration of the corresponding p serving as the probability of success. If these estimates of uncaptured individuals are added to the fixed number of captures, it provides an estimate of abundance with each posterior draw of capture probability and the distribution of abundances are statistically similar to those derived from N-augmentation.

To illustrate this point, we designed a small simulation study based on a simple closed model with a constant capture probability, p across three visits and a population size, N, of 1,000 individuals. We simulated 200 data sets based on either low ($p = 0.1$) or high ($p = 0.5$) capture probability and fit each data set using either a JD approach with N-augmentation or a JM approach with draws from the negative binomial distribution (JM-NB). We compared coverage of 50% and 95% confidence intervals as well as relative bias between the two estimation approaches. Both approaches had minimal positive relative bias (<0.4%) in the 100 simulations with a high capture probability. When capture probability was low, both approaches had modest positive relative bias, however the JM-NB approach was less biased than the N-augmentation approach (4.9% vs. 7.5%). Coverage of 50% and 95% confidence intervals was similar or slightly better for the JM-NB approach under both high and low capture probability scenarios. When capture probability was low, the JM-NB approach was 366 times faster at producing effective samples than the N-augmentation approach; however, when capture probability was high, the JM-NB approach was only 60 times faster.

In summary, the JM-NB approach is substantially faster than the popular N-augmentation approach and performed as well or better with respect to bias and coverage in our small simulation study. It can be extended to account for different priors on abundance and include individual heterogeneity (see King et al. 2016 for a detailed examination) and also easily extended to open models (Yackulic et al. 2018b) in which computation gains can be even more pronounced. We next reexamine the community occupancy model from application 3 to show how imperfect estimates of species presence can yield estimates of species richness that are statistically indistinguishable from estimates from a discrete community occupancy model.

Application 3 revisited

Community occupancy models typically estimate species richness for a given site and time by summing the discrete latent states for individual species. When discrete latent states are marginalized, it is possible to derive a point estimate of species richness by summing the probabilities of occurrence for individual species derived from one of the types of conditional probabilities. However, such an approach will underestimate variance in species richness (Fig. 4B). An alternative is to simulate a Bernoulli trial for each species based on its predicted probability of occurrence and sum these trials. While it seems curious to simulate discrete latent states after marginalizing them, this process is relatively quick since it does not lead to the autocorrelation in MCMC draws common to discrete latent state approaches and can be calculated as a generated quantity after parameters have been estimated. We derived estimates of species richness for application 3 for all sites and time periods using the JD, JM, and SM approaches as part of the earlier model fitting. For JM and SM approaches, probabilities of occurrence were based on forward conditional estimates and led to similar estimates of posterior means

and variances. Here, we focus on comparing estimates from JD and JM approaches. Mean estimates of first-order states are nearly identical in the final year (R^2 of 0.999 with minor discrepancies attributable to MCMC error) and highly correlated in earlier years (Fig. 4A). When estimates of first-order states from the JM version are summed, point estimates of species richness are similar to those derived from the JD version, but uncertainty is underestimated (Fig 4B). Both point estimates and variances on species richness are nearly identical provided binomial trials based on expected occupancy from the JM version are summed (Fig. 4C).

Auto-Covariates: Do We Lose Anything Through Marginalization?

The general concept that individual vital rates are influenced by second-order latent states is ubiquitous in ecological theory and thought to play an important role in many applied situations. For example, it is often hypothesized that recruitment and/or survival depend on population size, decreasing as the population increases and providing a mechanism for population regulation. Similarly, it is often hypothesized that rates of colonization or extinction of different habitat patches or territories will depend on the number of already occupied patches or territories in near proximity (i.e., autologistic effects). Autologistic effects can be based on the proportion or number of occupied patches/territories in some neighborhood or by weighting patches or territories by their distance and/or size. Failure to include population processes like density dependence and autologistic effects can lead to poor inferences and predictions, however, estimating the strength of these effects in field settings is not straightforward because estimates of vital rates and second-order states are uncertain and typically estimated jointly with non-negligible sampling correlations. In most circumstances, discrete latent state approaches provide unbiased estimates of these effects and it is unclear how to approach such effects if discrete latent states have been marginalized out to improve speed. In our final two applications, we compare parameter estimates (and speed) between JD, JM, and SM versions in which the marginalized versions use forward conditional estimates to calculate auto-covariates (an autologistic term in the fifth application and a density-dependent term in the final application). After discussing these applications, we consider other approaches that have been suggested for calculating second-order states for use as auto-covariates.

*Application 5: Two-species occupancy model with an autologistic effect applied to Northern Spotted Owls (*Strix occidentalis caurina*) and Barred Owls (*Strix varia*) in Tyee study area in Oregon, USA*

Northern Spotted Owls are threatened by both habitat loss and competition with invading Barred Owls (Yackulic et al. 2019). Two-species dynamic occupancy models have been used to quantify the impacts of competition and habitat on Northern spotted owl dynamics (Yackulic et al. 2014a, Dugger et al. 2016) and autologistic effects defining the whole study area as a neighborhood describe temporal patterns in barred owl colonization

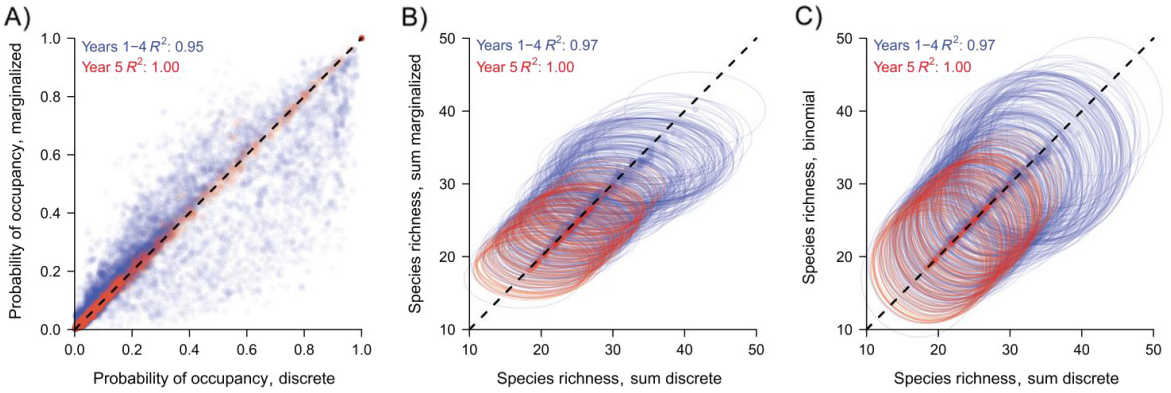

Fig. 4. Comparisons of occupancy and species richness estimates between marginalized and discrete versions of a community occupancy model. (A) Although a few of the 68,540 posterior mean occupancy estimates differ substantially between a discrete (*x*-axis) and marginalized (*y*-axis) version of a community occupancy model, estimates in the final year of the study are almost perfectly correlated (with slight discrepancies attributable to MCMC error) and estimates in the first four years of the study are highly correlated. (B) Summing marginalized estimates of occupancy (*y*-axis) from each posterior draw within sites to estimates species richness leads to average estimates that are highly correlated with species richness estimates from a discrete version (*x*-axis) in the first four years and almost perfectly correlated in the final year. However, this approach underestimates uncertainty in species richness estimates by 55% as indicated by the ovals centered on the mean estimates and given radii equivalent to the standard deviation of species richness estimates from the two versions. (C) Summing binomial draws based on each posterior occupancy estimate from a marginalized version leads to the same mean expectations and yields standard deviations in species richness that are very similar (4% greater) to those derived from a discrete approach as indicated by the nearly circular nature of the ovals. Dotted line in each plot indicates the 1:1 line.

and extinction well (Yackulic et al. 2012). Two-species dynamic occupancy models include parameters representing the initial occupancy of both species in the first season (year), the probability of local colonization and extinction for each species, and the probabilities of detection for each species. The data analyzed here were collected during 1990–2011 from 158 contiguous survey polygon (i.e., patches) that covered an area of approximately 1,000 km^2 in western Oregon, USA known as the Tyee study area. Occupancy for Northern Spotted Owls was defined in terms of territorial pairs, whereas occupancy of Barred Owls indicates the presence of one or more barred owls. Here, we fit different versions of the best model identified in Yackulic et al. (2014a) and refer readers to this publication for more details. Specifically, we fit a JD version of the model, and compare it to JM and SM versions in which the autologistic covariate is calculated based on forward conditional estimates of Barred Owl occupancy. We also compare this to previous published estimates based on a closed conditional, maximum likelihood approach to estimating barred owl occupancy. Last, we fit a version of the SM model with spatial and temporal additive random effects to further illustrate how pseudo-R^2 can be calculated.

Estimates and standard errors of all parameters were indistinguishable among JD, JM, and SM versions. For example, the estimates of the autologistic effect on colonization were 4.2 with a standard error of 0.6 for all three versions. This differed slightly from a past estimate of 4.8 with a standard error of 0.5 reported by Yackulic et al. (2014a). While the parameter estimates for JM and SM autologistic effects were based on forward conditional estimates of occupancy, past estimates used closed conditional estimates of occupancy, providing weak evidence that forward conditional estimates may outperform closed conditional estimates as auto-covariates. The SMRE version led to more substantial differences in parameters. In particular, the size of the autologistic and habitat effects on both Barred Owl colonization and extinction increased.

JD and JM versions took a similar amount of time per 100 effective samples, though on average JM was slightly slower (304 minutes vs. 270 minutes). The SM was substantially faster, requiring an average of 16 minutes per 100 effective samples. The SMRE version was slower than the SM version (37 minutes per 100 effective samples), but still much quicker than either of the JAGS-based versions that lacked random effects. Inclusion of additive site and year random effects alongside fixed effects (i.e., habitat, competition, and autologistic covariates) allowed us to calculate multi-level pseudo-R^2 values, which quantified the explained spatial and temporal variance in colonization and extinction of Barred and Northern Spotted Owls. For Barred Owls, covariates explained a high proportion of temporal variation in colonization (pseudo-$R^2 = 0.94$) and extinction (pseudo-$R^2 = 0.61$), but less spatial variation (0.10 and 0.33 of spatial variation in colonization and extinction explained), suggesting that additional covariates may be required to explain Barred Owl dynamics. For Northern Spotted Owls, 0.49 of spatial variation in colonization was explained, while all other pseudo-R^2's were less than 0.4, suggesting that our covariates are not addressing important drivers of variation in Northern Spotted Owl dynamics. This example demonstrates the utility of the pseudo-R^2 approach to identify significant drivers of ecological dynamics and shortfalls of existing models.

Application 6: N-occupancy model applied to barred owls in the Tyee study area in Oregon, USA

As the last example illustrates, Barred Owl detection–nondetection data can be analyzed in an occupancy framework to provide inferences about territorial colonization and extinction, however, managers are increasingly interested in understanding barred owl demographic parameters. Existing knowledge of barred owl demography is based primarily on an intensive study in the Coastal Ranges study area in Western Oregon (Wiens et al. 2014). Rossman et al. (2016) developed an *N*-occupancy model to infer demographics from the same Barred Owl detection-nondetection data analyzed in the last application (but with four additional years of data). The *N*-occupancy model assumes an underlying demographic model in each site with parameters dictating initial site abundances, apparent survival, and gains.

In the demographic model, gains were modeled as a function of mean density in the study area (second-order state) and apparent survival varied with a habitat covariate. In the marginalized versions, the mean density autocovariate was calculated based on forward conditional estimates of Barred Owl site abundances. The latent abundances in each site were linked to site detection via the Royle-Nichols relationship (Royle and Nichols 2003). Marginalizing this model requires setting an upper limit for site abundance, Ninf, and then keeping track of the probability of all site abundances less than Ninf (including zero) for each site and year. If Ninf is set too low, estimates will be biased; however, the model will needlessly track site abundances slowing computation time if it is set too high (bias when Ninf is too small has been observed for *N*-mixture models, a similar class of models; see Couturier et al. 2013). We fit JD and SM versions of this model as well as a maximum likelihood version. The maximum likelihood was used to test different values of Ninf and determine that a value of 10 led to indistinguishable parameter estimates when compared to higher values of Ninf. We also used the maximum likelihood version to compare the Akaike Information Criterion (AIC) of the *N*-occupancy model to a standard dynamic occupancy model fit with same data and a similar structure (i.e., habitat covariate on extinction and autologistic covariate on colonization).

JD and SM versions of the model led to parameter estimates whose credible intervals largely overlapped. In particular, the parameter relating mean density in a given year to gains in the next year was estimated as 0.8

with a standard error of 0.1 by both versions. The SM version estimated the habitat effect as 50% larger in magnitude than the JD version, but with wider and overlapping 95% credible intervals (0.4–2.2 vs. 0.3–1.3). The JD version took an average of over 8 h (513 minutes) per 100 effective samples, while the SM version took an average of 18 minutes. The maximum likelihood version yielded similar estimates to the SM approach and took less than five minutes. AIC strongly favored the *N*-occupancy model over a standard occupancy model with the same number of parameters and covariates (ΔAIC = 61).

Are perfect estimates of conditional probabilities necessary?

In the fifth and sixth applications, we used forward conditional estimates of first-order states to generate time-varying estimates of a second-order state (neighborhood occupancy and density) and found minimal differences from discrete latent state approaches (i.e., fully conditional approaches). Yackulic et al. (2012) analyzed the Barred Owl data presented in application 5 using a closed conditional approach to estimate occupancy and found small differences from a discrete latent state approach. Yackulic et al. (2012) also performed a simulation study and found minimal to no bias when a closed conditional approach was used. In another study, Yackulic et al. (2018b) used the negative binomial approach to estimating abundance (described above), which was then used both as a covariate in a model of survival for the same species (density dependence) and as a covariate in models of survival and capture probability of a competing species (interspecific interactions). Simulation studies suggested minimal bias using these approaches as well. Taken together, available evidence suggests that using closed conditional or forward conditional approaches to generate estimates of second-order states has a minimal impact on analyses, with weak evidence that forward conditional approaches are preferable to closed conditional approaches.

Since fully conditional estimates are the same as those derived from discrete latent states, it is logical that they might provide as good, or better, inferences than the imperfect estimators discussed in the last paragraph. However, trying to use fully conditional estimates in an autologistic model leads to a series of circular references as fully conditional approaches require estimates of the likelihood of full capture, or detection histories, and these depend on estimates of parameters, such as colonization, or survival, which in an autologistic or density-dependent, model require the fully conditional estimates of occupancy. This situation differs from a forward conditional approach, where neighborhood occupancy can be estimated for time t and used to predict colonization in the interval between times t and $t + 1$, which can itself be used to estimate neighborhood occupancy in time $t + 1$. Hall et al. (2018) proposed an approach in a maximum likelihood framework to iteratively arrive at fully conditional estimates of occupancy for an autologistic analysis by calculating fully conditional estimates based on an initial set of parameters, using these estimates as a covariate and recalculating parameters, and then using the new parameters to estimate and updated set of fully conditional estimates. Hall et al. (2018) advocated this approach and argued that other approaches (i.e., closed conditional or forward conditional), while computationally efficient, were "insufficient" because they did not use all the information in a detection history (i.e., because they are imperfect estimators). Unfortunately, they did not include a simulation study comparing their approach to one of the imperfect approaches. To address this gap in our understanding, we ran a simulation study comparing the iterative fully conditional approach to a forward conditional approach.

A comparison of an iterative fully conditional approach to autologistic modeling and the forward conditional approach

For all simulations, we assumed that extinction and detection probabilities were 0.5 and that colonization was given by a logit-linear equation with an intercept of −1 and a slope of 2 (which was multiplied by the autologistic covariate). We assumed 100 sites, six seasons (five intervals), and two visits per season. We defined the autologistic covariate based on a limited neighborhood consisting of the two sites before and after a site when they were ordered as simulated. We considered two sets of simulated data. In the first set, initial occupancy for 100 sites was randomly simulated based on a binomial trial with a 0.5 probability of success, while in the second set the first 50 sites were assumed to be unoccupied and the other 50 sites were assumed to be occupied. We ran this second set, because we hypothesized the fully conditional approach might perform better when there was unmodeled spatial autocorrelation. For each set, we simulated 1,000 data sets and fit each data set using both modeling approaches and a maximum likelihood framework. We compared the bias and coverage of parameter estimates from the two approaches.

We found little evidence to suggest an iterative fully conditional approach was an improvement over a forward conditional approach. Both approaches had minimal bias in all parameters when initial occupancy was random. Similarly, when initial occupancy was random, coverage of 50% and 95% confidence intervals was adequate and did not differ appreciably between approaches. When initial occupancy was clustered, we found similar levels of bias for both approaches in estimates of the intercept and slope of colonization (Fig. 5). Coverage of 50% confidence intervals on the intercept was only 44.8% for the fully conditional approach and 43.6% for the forward conditional approach. However, coverage was adequate for all other parameters at the

50% level and for all parameters at the 95% level and did not differ between approaches. While there may well be circumstances in which imperfect estimators of first-order states, such as unconditional and forward conditional estimates, lead to biases, they have yet to be identified and analyses to date suggest that these approaches can be used to calculate auto-covariates and provide reasonable inferences.

Summary

Marginalization (and using Stan) leads to much faster estimates

In all six applications we considered, the SM version of a model provided similar estimates to the JD version, while requiring a fraction of the time to provide these estimates (Fig. 6) with one exception. In the humpback chub multistate mark–recapture application (application 2), we cannot confirm that SM and JD parameter estimates are the same since the JD version was still far from converging after 8.3 d even though the SM versions had converged and produced more than 500 effective samples of all parameters in less than two hours. For this application and the first application, the time to reach a minimum of 100 effective samples declined by at least three orders of magnitude following marginalization. Speed gains were more modest in the other four applications, in which the SM version was only 5–25 times faster than the JD version. Comparing JD and SM versions encompasses two important changes, marginalization and a change in software (Appendix S1: Fig. S2). If we focus on the five applications for which we fit JD and JM versions, we can isolate the significance of marginalization alone. In applications 1 and 2, the JM version was approximately two orders of magnitude faster than the JD version. In application 3, the JM code was ~20 times faster, while for applications 4 and 5, the two approaches took roughly the same amount of time.

While we focused on Stan and different programs that use the BUGS language, our guide to marginalization should be useful to users applying marginalization in other software both now and in the future as other software is developed. For example, we have seen similar gains in speed from marginalizing code for samplers written directly in program R. In addition, programs like Nimble support marginalization, but also provide flexibility in sampling methodologies and are constantly improving (de Valpine et al. 2017). While Nimble does not yet include the Hamiltonian Monte Carlo algorithms that underlie Stan, it does include other features such as dynamic block sampling that can improve speed significantly for some models (Turek et al. 2016), and may soon include the Hamiltonian Monte Carlo

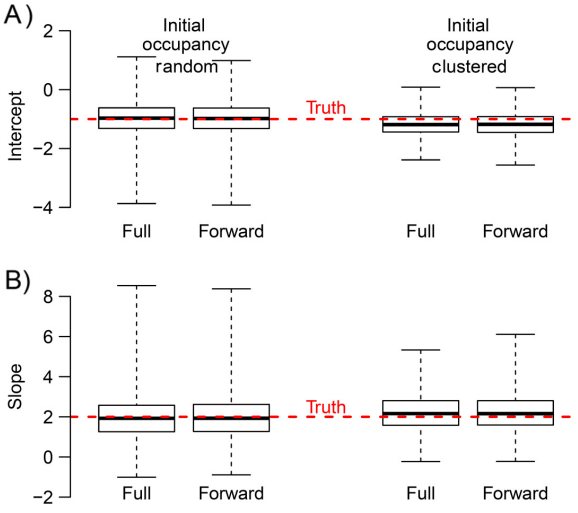

Fig. 5. Models that use full or forward conditional approaches to estimate an autologistic covariate on colonization provide similar estimates of the (A) intercept and (B) slope. When initial occupancy is random neither approach shows bias and when initial occupancy is clustered both approaches show similar levels of bias, box plots are based on fitted estimates from 1,000 simulated data sets, with box edges representing the inner quartiles, the black line representing the median and whiskers covering the full range of variation. Dotted red lines labeled truth show the value used to simulate the data.

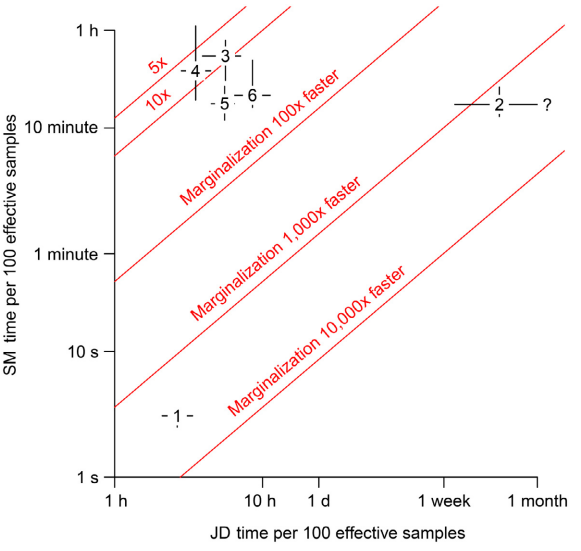

Fig. 6. Comparison of the time required to reach at least 100 effective samples for all parameters (i.e., speed) when data from the six applications were analyzed using discrete versions of code implemented in JAGS (JD) or marginalized versions of code implemented in Stan (SM). Black numbers indicate the application number and error bars cover the range of speeds observed when each version was fit 10 times (for some applications, error bars are not distinguishable if range of speeds is less than size of the application number text). Red lines indicate ratios of speed between the two approaches. For application 2, the JD version had less than 10 effective samples after 8.3 d, which suggests it might have taken more than a month to reach 100 effective samples; however, uncertainty in speed is indicated by a question mark.

algorithm (P. de Valpine, *personal communication*). In all our examples we used options in the R programs that run Stan, JAGS, BUGS, and OpenBUGS that allow for different threads to be analyzed in parallel, another approach for speeding up analyses. In the future, additional gains in speed may come from parallelization within threads (i.e., splitting up likelihood calculations within a MCMC iteration), a process that is likely to be much more straightforward with marginalized code.

While most applications we considered took hours as JD versions, we see no reasons why similar, or better gains in speed would not occur if we applied marginalization to discrete models that currently take days or weeks. Furthermore, while the processors on the server we used are not particularly fast, we have found similar relative differences on other faster systems, suggesting that marginalization remains useful for most analyses given typical computer specifications. Lastly, we make no claims that any version is optimized, and we expect that others will be able to improve the speed of marginalized code presented here. The practical gains from increases in speed by the orders of magnitude we report should not be understated. Widespread application of marginalization in Bayesian population models should allow more thorough simulation studies, more emphasis on comparison of alternative model structures, and quicker learning.

Costs associated with marginalization are minimal

Users already familiar with marginalization often focus on perceived issues relating to recovery of first-and second-order discrete latent states. In *Recovering discrete latent states: issues and terms*, we described (and named) different approaches for recovering first-and second-order discrete latent states. Fully conditional estimates of first-order states are the product of discrete latent state approaches and are the best estimates. However, in most applications, inferences are focused on parameters, which are typically unaffected by estimates of first-order discrete latent states, or second-order states. Estimates of second-order discrete latent states can often be derived from imperfect estimates of first-order discrete latent states or by ignoring first-order discrete latent states all together, and these estimates are indistinguishable from estimates based on discrete latent state approaches (Fig. 4).

In *Auto-covariates: do we lose anything through marginalization?*, we further showed that models that include auto-covariates are also not appreciably affected by the use of imperfect forward conditional estimates of states in the place of fully conditional estimates (Fig. 5). Simulations and applications suggested inferences from these two approaches are the same and that models using forward conditional estimates are much quicker. While it is plausible that for some models or data, differences in inferences may occur due to marginalization, we struggled to identify these situations.

Understanding how marginalization works shrinks the divide between Bayesian and maximum likelihood approaches to population models

While we recognize philosophical reasons for favoring Bayesian or maximum likelihood approaches, our experience is that most ecologists choose an approach for a given analysis based on pragmatic concerns. Models that assume an underlying latent site abundance have often been presented in discrete versions and thus only applied using Bayesian approaches. In application 6, we fit a maximum likelihood approach to an *N*-occupancy model alongside SM and JD version and showed it was much quicker than even the SM version. Development of a maximum likelihood version allowed us to calculate the AIC of this model and compare it to that of a comparable occupancy model in less time than it took to run even the SM version. In another example, researchers have recently determined ways to fit open spatial capture–recapture models using marginalization and hidden Markov modeling approaches (Glennie et al. 2019). Researchers are also working on marginalizing continuous latent states in state-space models of abundance (Besbeas and Morgan 2020), which could eventually improve speed and accuracy in joint abundance models increasingly used in community ecology (Warton et al. 2015).

In *To Bayes or not to Bayes: a pragmatic response*, we emphasized a few of the reasons that might lead an ecologist to favor a Bayesian approach for a particular analysis. Application 4 illustrated how priors can improve inferences when data are sparse (Fig. 3), while addition of random effects to application 2 allowed us to infer temporal variation in movement that was otherwise obscured (Fig. 2). Random effects can be implemented in maximum likelihood using programs such as MARK, however, Bayesian approaches become preferable as random effects are added to multiple parameters (or when users wish to include spatially or temporally structured random effects) or in the case of community occupancy models where random effects are core component of the model structure. Inclusion of random effects also allows user to calculate pseudo-R^2 for different parameters, which improved understanding in the two applications with covariates. Applying marginalization allows users to take advantage of all these potential benefits of a Bayesian approach while minimizing the speed that is sacrificed using these methods.

Acknowledgments

Jim Nichols, Mary Conners, and Scott VanderKooi provided valuable feedback on earlier drafts of this manuscript. The various data sets analyzed in this manuscript represent years of hard work by various dedicated field crews and we are grateful to all the crews that contributed to these various efforts. Data analyzed in applications 1, 2, and 4 were funded by the Glen Canyon Dam Adaptive Management Program and collected by researchers from various agencies with special thanks to Michael Yard and Josh Korman, who led multiple mark–recapture efforts in the mainstem Colorado River, U.S. Fish and

Wildlife Service, who led spring and fall mark–recapture efforts in the Little Colorado River, Bill Pine and the Near Shore Ecology project who led sampling in the Colorado River study reach from 2009 through 2011, and Arizona Game and Fish who collected brown trout catch data in the Glen Canyon reach. Logistic support and technical assistance for the data analyzed in application 3 were provided by the Coronado National Forest, the Nature Conservancy, Fort Huachuca, Santa Rita Experimental Station, and Audubon Research Ranch. Support for the data analyzed in applications 5 and 6 was provided by the U.S. Forest Service and the Bureau of Land Management. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Literature Cited

Arnason, N. A. 1973. The estimation of population size, migration rates and survival in a stratified population. Researches on Population Ecology 15:1–8.

Barry, S. C., S. P. Brooks, E. A. Catchpole, and B. J. T. Morgan. 2003. The analysis of ring-recovery data using random effects. Biometrics 59:54–65.

Besbeas, P., and B. J. T. Morgan. 2020. A general framework for modeling population abundance data. Biometrics 76:281–292.

Betancourt, M.2017. A conceptual introduction to Hamiltonian Monte Carlo. arXiv preprint arXiv:1701.02434

Bonner, S., and M. Schofield. 2014. MC (MC) MC: exploring Monte Carlo integration within MCMC for mark–recapture models with individual covariates. Methods in Ecology and Evolution 5:1305–1315.

Brownie, C., J. E. Hines, J. D. Nichols, K. H. Pollock, and J. B. Hestbeck. 1993. Capture-recapture studies for multiple strata including non-Markovian transitions. Biometrics 49:1173–1187.

Burnham, K. P., and G. C. White. 2002. Evaluation of some random effects methodology applicable to bird ringing data. Journal of Applied Statistics 29:245–264.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. 2017. Stan: A probabilistic programming language. Journal of Statistical Software 76:32.

Cormack, R. 1964. Estimates of survival from the sighting of marked animals. Biometrika 51:429–438.

Couturier, T., M. Cheylan, A. Bertolero, G. Astruc, and A. Besnard. 2013. Estimating abundance and population trends when detection is low and highly variable: a comparison of three methods for the Hermann's tortoise. Journal of Wildlife Management 77:454–462.

de Valpine, P., D. Turek, C. J. Paciorek, C. Anderson-Bergman, D. T. Lang, and R. Bodik. 2017. Programming with models: writing statistical algorithms for general model structures with NIMBLE. Journal of Computational and Graphical Statistics 26:403–413.

Dorazio, R. M., J. A. Royle, B. Söderström, and A. Glimskär. 2006. Estimating species richness and accumulation by modeling species occurrence and detectability. Ecology 87:842–854.

Dugger, K. M., et al. 2016. The effects of habitat, climate, and Barred Owls on long-term demography of Northern Spotted Owls. Condor 118:57–116.

Garrard, G. E., M. A. McCarthy, P. A. Vesk, J. Q. Radford, and A. F. Bennett. 2012. A predictive model of avian natal dispersal distance provides prior information for investigating response to landscape change. Journal of Animal Ecology 81:14–23.

Gelman, A., and J. Hill 2007. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, New York, New York, USA.

Gelman, A., and L. Pardoe. 2006. Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. Technometrics 48:241–251.

Glennie, R., D. L. Borchers, M. Murchie, B. J. Harmsen, and R. J. Foster. 2019. Open population maximum likelihood spatial capture-recapture. Biometrics 75:1345–1355.

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

Guillera-Arroita, G., M. Kéry, and J. J. Lahoz-Monfort. 2019. Inferring species richness using multispecies occupancy modeling: estimation performance and interpretation. Ecology and Evolution 9:780–792.

Hall, L. A., N. D. Van Schmidt, and S. R. Beissinger. 2018. Validating dispersal distances inferred from autoregressive occupancy models with genetic parentage assignments. Journal of Animal Ecology 87:691–702.

Itô, H.2017. Partial Stan translation of Bayesian population analysis. https://github.com/stan-dev/example-models/tree/master/BPA

Kéry, M. 2010. Introduction to WinBUGS for ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses. Academic Press, Cambridge, Massachusetts, USA.

Kéry, M., and M. Schaub 2011. Bayesian population analysis using WinBUGS: a hierarchical perspective. Academic Press, Cambridge, Massachusetts, USA.

King, R., and S. Brooks. 2001. On the Bayesian analysis of population size. Biometrika 88:317–336.

King, R., B. T. McClintock, D. Kidney, and D. Borchers. 2016. Capture-recapture abundance estimation using a semi-complete data likelihood approach. Annals of Applied Statistics 10:264–285.

Korman, J., and M. Yard. 2017. Effects of environmental covariates and density on the catchability of fish populations and interpretation of catch per unit effort trends. Fisheries Research 189:18–34.

Korman, J., M. D. Yard, and C. B. Yackulic. 2015. Factors controlling the abundance of rainbow trout in the Colorado River in Grand Canyon in a reach utilized by endangered humpback chub. Canadian Journal of Fisheries and Aquatic Sciences 73:105–124.

Lebreton, J.-D., J. D. Nichols, R. J. Barker, R. Pradel, and J. A. Spendelow. 2009. Modeling individual animal histories with multistate capture-recapture models. Pages 87–173 *in* H. Caswell, editor. Advances in ecological research. Academic Press, Burlington, New Jersey, USA.

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

Lorenzen, K. 2000. Allometry of natural mortality as a basis for assessing optimal release size in fish-stocking programmes. Canadian Journal of Fisheries and Aquatic Sciences 57:2374–2381.

Lunn, D., D. Spiegelhalter, A. Thomas, and N. Best. 2009. The BUGS project: evolution, critique and future directions. Statistics in Medicine 28:3049–3067.

Lunn, D. J., A. Thomas, N. Best, and D. Spiegelhalter. 2000. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 10:325–337.

Monnahan, C. C., J. T. Thorson, and T. A. Branch. 2017. Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. Methods in Ecology and Evolution 8:339–348.

Northrup, J. M., and B. D. Gerber. 2018. A comment on priors for Bayesian occupancy models. PLoS ONE 13:e0192819.

Plummer, M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. Pages 1–10 *in* Proceedings of the 3rd International Workshop on Distributed Statistical Computing, Vol. 124.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: convergence diagnosis and output analysis for MCMC. R News 6:7–11.

Rankin, R. W., K. E. Nicholson, S. J. Allen, M. Krützen, L. Bejder, and K. H. Pollock. 2016. A full-capture hierarchical Bayesian model of Pollock's closed robust design and application to Dolphins. Frontiers in Marine Science 3:1–25.

Rossman, S., C. B. Yackulic, S. P. Saunders, J. Reid, R. Davis, and E. F. Zipkin. 2016. Dynamic N-occupancy models: estimating demographic rates and local abundance from detection-nondetection data. Ecology 97:3300–3307.

Royle, J. A., and R. M. Dorazio 2008. Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities. Academic Press, San Diego, California, USA.

Royle, J. A., and R. M. Dorazio. 2012. Parameter-expanded data augmentation for Bayesian analysis of capture–recapture models. Journal of Ornithology 152:521–537.

Royle, J. A., and W. A. Link. 2002. Random effects and shrinkage estimation in capture-recapture models. Journal of Applied Statistics 29:329–351.

Royle, J. A., and J. D. Nichols. 2003. Estimating abundance from repeated presence-absence data or point counts. Ecology 84:777–790.

Runge, M. C., C. B. Yackulic, L. S. Bair, T. A. Kennedy, R. A. Valdez, C. Ellsworth, J. L. Kershner, R. S. Rogers, M. A. Trammell, and K. L. Young. 2018. Brown trout in the Lees Ferry reach of the Colorado River—Evaluation of causal hypotheses and potential interventions. U.S. Geological Survey Open-File Report 2018–1069. Reston, Virginia, USA, 83 p.

Russell, R. E., J. A. Royle, V. A. Saab, J. F. Lehmkuhl, W. M. Block, and J. R. Sauer. 2009. Modeling the effects of environmental disturbance on wildlife communities: avian responses to prescribed fire. Ecological Applications 19:1253–1263.

Sanderlin, J. S., W. M. Block, and J. L. Ganey. 2014. Optimizing study design for multi-species avian monitoring programmes. Journal of Applied Ecology 51:860–870.

Sanderlin, J. S., W. M. Block, J. L. Ganey, and J. M. Iniguez 2013. Preliminary assessment of species richness and avian community dynamics in the Madrean Sky Islands, Arizona. Pages 180–190 *in* G. J. Gottfried, P. F. Ffolliott, B. S. Gebow, L. G. Eskew, and L. C.Collins, editors. Merging science and management in a rapidly changing world. RMRS-P-67. US Forest Service, Fort Collins, Colorado, USA.

Schofield, M. R., and R. J. Barker. 2008. A unified capture-recapture framework. Journal of Agricultural, Biological, and Environmental Statistics 13:458.

Schwarz, C. J., J. F. Schweigert, and A. N. Arnason. 1993. Estimating migration rates using tag-recovery data. Biometrics 49:177–193.

Seaman, J. W., J. W. Seaman, and J. D. Stamey. 2012. Hidden dangers of specifying noninformative priors. American Statistician 66:77–84.

Tanner, M. A., and W. H. Wong. 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82:528–540.

Turek, D., P. de Valpine, and C. J. Paciorek. 2016. Efficient Markov chain Monte Carlo sampling for hierarchical hidden Markov models. Environmental and Ecological Statistics 23:549–564.

Warton, D. I., F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F. K. C. Hui. 2015. So many variables: joint modeling in community ecology. Trends in Ecology & Evolution 30:766–779.

Wiens, J. D., R. G. Anthony, and E. D. Forsman. 2014. Competitive interactions and resource partitioning between northern spotted owls and barred owls in western Oregon. Wildlife Monographs 185:1–50.

Yackulic, C. B. 2018. Humpback Chub (*Gila cypha*) and Rainbow Trout (*Oncorhynchus mykiss*) Joint Mark-Recapture Data and Model, Colorado River, Arizona. U.S. Geological Survey data release. https://doi.org/10.5066/F7ZC81T9

Yackulic, C. B., M. Dzul, J. A. Reid, J. S. Sanderlin, W. M. Block, J. L. Ganey, M. Dodrill, and M. D. Yard. 2020. Marginalizing Bayesian population models—data for examples in the Grand Canyon region, Southeastern Arizona, and western Oregon, USA—1990–2015. U.S. Geological Survey data release. https://doi.org/10.5066/P9JN5C0L

Yackulic, C. B., J. Korman, and L. Coggins. 2018*a*. Population dynamics of humpback chub, rainbow trout and brown trout in the Colorado River in its Grand Canyon Reach: modelling code and input data. U.S. Geological Survey data release. https://doi.org/10.5066/F7FN15HC

Yackulic, C. B., J. Korman, M. D. Yard, and M. Dzul. 2018*b*. Inferring species interactions through joint mark–recapture analysis. Ecology 99:812–821.

Yackulic, C. B., J. Reid, R. Davis, J. E. Hines, J. D. Nichols, and E. Forsman. 2012. Neighborhood and habitat effects on vital rates: expansion of the Barred Owl in the Oregon Coast Ranges. Ecology 93:1953–1966.

Yackulic, C. B., J. Reid, J. D. Nichols, J. E. Hines, R. Davis, and E. Forsman. 2014*a*. The roles of competition and habitat in the dynamics of populations and species distributions. Ecology 95:265–279.

Yackulic, C. B., M. D. Yard, J. Korman, and D. R. Van Haverbeke. 2014*b*. A quantitative life history of endangered humpback chub that spawn in the Little Colorado River: variation in movement, growth, and survival. Ecology and Evolution 4:1006–1018.

Yackulic, C. B., et al. 2019. The past and future roles of competition and habitat in the range-wide occupancy dynamics of Northern Spotted Owls. Ecological Applications 29: e01861.

Supporting Information

Additional supporting information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/eap.2112/full

Data Availability

Additional data are available on ScienceBase: https://doi.org/10.5066/P9JN5C0L