Efficient Bayesian Estimation for Open Population Capture-Recapture Models Without Data Augmentation

Devin S. Johnson
Pacific Islands Fisheries Science Center,
National Marine Fisheries Service, NOAA,
Honolulu, Hawai'i, USA
email: devin.johnson@noaa.gov

Shelbie K. Ishimaru
Hawai'i Institute of Marine Biology,
Cooperative Institute of Marine and Atmospheric Research,
University of Hawai'i at Mānoa,
Honolulu, Hawai'i, USA

Janelle J. Gardner
Pacific Islands Fisheries Science Center,
National Marine Fisheries Service, NOAA,
Honolulu, Hawai'i, USA

July 7, 2026

Abstract

1. Bayesian estimation of abundance with capture-recapture data has been dominated for nearly 20 years by the parameter-expansion data-augmentation (PX-DA) approach. The PX-DA approach expands the parameter set to include the latent true states of the individuals. PX-DA allows straightforward coding of models in MCMC software such as `JAGS` (Just Another Gibbs Sampler) or `nimble`, however, this approach can be computationally demanding for large or low detectability populations.

2. We develop a collapsed Gibbs sampler version of the PX-DA approach to reduce computational burden when using Markov Chain Monte Carlo (MCMC) to fit Jolly-Seber (JS) models. Using the collapsed sampler, no parameter expansion or data augmentation is needed. Two of the main computational benefits are: (1) the MCMC only needs to iterate over unique capture-histories and (2) the MCMC can be divided into 2 stages where the second stage sampler can be run in parallel. We provide an `R` package `nimbleJSextras` that uses the functionality of the `nimble` and `nimbleEcology` packages to perform MCMC analysis for JS models with the same ease for practitioners the PX-DA approach.

3. Using the `nimbleJSextras` package, we analyzed two real data sets to illustrate the method. First we analyzed the classic dipper data using *per capita* recruitment parameterization. We compared the collapsed sampler to PX-DA. In the second example we analyzed a data set of 5,271 female nesting green turtles in the Northwest Hawaiian Islands over a 44 year period that contained time by individual behavioral effects induced by capture procedures.

4. In the dipper analysis the relative speed of the MCMC compared to the PX-DA sampler depended on the number of augmented individuals, slower for 300 augmented individuals and faster for 1,000 individuals. Convergence metrics showed slightly quicker convergence of the collapsed MCMC than the PX-DA sampler. The full 2-stage collapsed sampler ran in only 1/10 the amount of time needed for the PX-DA sampler. The more complex green turtle data finished in < 3.5 hours illustrating that relatively large, long-term data can be analyzed efficiently even with a complex capture probability model using the 2-stage collapsed sampler. Moreover, trial analyses are not necessary to determine an adequate number of augmented individuals.

Key words: Abundance estimation, Bayesian, capture-recapture, collapsed Gibbs sampler, Markov Chain Monte Carlo, parallel computing

1 Introduction

For over 60 years, open-population capture–recapture models, particularly those in the Jolly–Seber (JS) family (Jolly, 1965; Seber, 1965), have been fundamental tools for estimating demographic rates such as survival, recruitment, and population size in open systems where individuals can enter or leave the population between sampling occasions. Classical maximum-likelihood estimators and their extensions (Schwarz and Arnason, 1996) have long provided practical means for inference, yet they are often limited in their capacity to accommodate hierarchical structures, covariate effects, or complex dependence among parameters. Bayesian approaches have addressed many of these limitations, enabling flexible modeling and coherent quantification of uncertainty (Royle and Dorazio, 2008).

Adoption of Bayesian inference for JS models was initially hindered in regular use because standard Bayesian Markov Chain Monte Carlo (MCMC) software at the time, such as `WinBUGS` (Lunn et al., 2000) or `JAGS` (Plummer, 2003) did not have the capability to calculate the JS likelihood within the MCMC sampler. Therefore, practitioners would have to code the JS likelihood as well as bespoke MCMC algorithms for Bayesian estimation of JS parameters. The groundbreaking papers of Durban and Elston (2005) and Royle et al. (2007) introduced the concept of a population of augmented individuals which may or may not be part of the true population of interest, depending on whether or not they entered based on an probabilistic binary indicator of inclusion. This parameter expansion data augmentation (PX-DA) specification was formalized for JS models by Royle and Dorazio (2012). By introducing latent inclusion indicators as part of the hierarchical model, PX-DA allows MCMC sampling in a fixed-dimensional parameter space, where abundance is simply the sum of the latent inclusion indicators and not a model parameter. This breakthrough allowed users to construct a JS model hierarchically in `WinBUGS` or `JAGS` using simple, standard distributions. Due to the ease in which JS models could be fit with PX-DA methods, regular use of Bayesian methods for estimating abundance from capture-recapture increased dramatically and augmentation methods became the dominant paradigm for abundance estimation in a Bayesian setting. The PX-DA approach opened many doors for fitting models that cannot be easily fit with maximum likelihood methods, namely those including individual heterogeneity components (e.g., Durban and Elston 2005 and Royle and Dorazio 2008, Ch. 10) and other hierarchical specifications for time-indexed effects.

Although data augmentation was instrumental in ushering in a new era a Bayesian JS modeling, the convenience comes at a cost. The approach can be computationally inefficient for analyses requiring a large amount of augmentation (Royle et al., 2007) and is prone to slow mixing (Schofield and Barker, 2008). For example, if the probability of capture is

relatively low and the population is large, many augmented individuals need to be added to the data and states for every individual must be updated even when their current inclusion indicator does not indicate they are part of the actual population. In addition, Link (2013) notes some questionable properties of PX-DA method relative to the discrete uniform prior induced by the parameter expansion. To alleviate some of these issues and speed computation Yackulic et al. (2020) and Saracco and Yackulic (2023) proposed the "marginalized" capture-recapture models, which formulates capture-recapture models as a Hidden Markov Model (HMM; Zucchini et al. 2016). The HMM forward algorithm allows one to marginalize over latent states so that they need not be sampled as part of the MCMC. This increases speed and often improves MCMC chain mixing. However, the user still augments data for undetected individuals and uses the MCMC to marginalize over those not included in the population. In this paper, we introduce a fully marginalized Bayesian formulation of the JS model that eliminates the need for any data augmentation. Our approach integrates analytically over both the unobserved individual encounter histories and the latent states of all individuals using a conditional HMM formulation of JS family models. We propose using the exact JS likelihood within a *collapsed* Gibbs sampler (Liu, 1994) to estimate parameters and derived quantities of JS models.

Collapsing a Gibbs sampler is a method to improve MCMC convergence over standard Gibbs samplers. Liu (1994) provides a mathematical proof which shows that the rate of convergence for a collapsed Gibbs sampler is at least as good, but often better, than the standard Gibbs sampler for any particular model. Unfortunately, we cannot directly use the theoretical results to prove that the collapsed Gibbs sampler for JS models will always converge at a faster rate because neither the PX-DA sampler nor the collapsed version presented here use strictly Gibbs updates. That is, at least some, if not all, of the parameters are updated using Metropolis-within-Gibbs updates because the full conditional distributions are not standard. However, Van Dyk and Jiao (2015) note that MCMC convergence for collapsed samplers using Metropolis updates can still offer improvement but it may not be as good as strictly Gibbs updates.

The overall goal of this paper is to develop and demonstrate the collapsed approach for JS models using the MCMC software in the `R` package `nimble` (de Valpine et al., 2017) to illustrate that PX-DA methods are not necessary and JS models can be fit without data augmentation in regular practice. The `nimble` package allows users to create their own functions which can be compiled and included within MCMC model statements. Using this capability, as well as the HMM likelihoods within the `R` package `nimbleEcology` (Goldstein et al., 2024) we developed an extension package, `nimbleJSextras`* that included JS

*Available at `https://github.com/dsjohnson/nimbleJSextras`

likelihood and other functions for using the collapsed Gibbs sampler. The benefits of this approach include the fact it is not necessary to specify a fixed number of augmented individuals that were not captured. Thus, practitioners do not need to worry about possible biases induced by adding too few individuals (Link, 2013). Moreover, because augmented individuals are not used, the likelihood scales with the number of unique capture-histories, making it more efficient for large samples and populations. Finally, we can use the likelihood factoring approach of King et al. (2016) and Hooten et al. (2023) to split the sampler into two stages where the second stage abundance sample can be executed in parallel. We demonstrate its use and benefits through analysis the classic European dipper (*Cinclus cinclus*) data as well as more complex breeding return-time JS model for green sea turtles (*Chelonia mydas*). The analysis and accompanying code illustrate that augmentation is not necessary and the fully marginal collapsed Gibbs sampling approach can be used by practitioners just as easily as PX-DA methods.

2 The Jolly-Seber class of models

Here we describe the class of Jolly-Seber models as we will use them to develop the collapsed Gibbs sampler methods. All of the variations we consider here have been previously described in the literature, so there are no new model specifications herein, but we are using a general form to include as many different JS aspects for open populations as possible.

In a JS process the population is described as N individuals that are available for capture in at least one capture occasion during the study period. If the study is relatively long-term, this may not be very useful, so, the occasion specific abundance, N_t, is often desired instead, where N_t is the size of the population available for capture on occasion $t = 1, \ldots, K$. Individuals become available and unavailable through an entry and exit process where individuals can be in 1 of 3 states: "pre-entry" (pre-birth), "available", and "post-exit" (post-death). For this paper we expand the set of states so that we may include general multistate models as well. Therefore, an individual can be in one of J states, $s_1, \ldots, s_J$ where s_1 is the pre-entry state and s_J is the post-exit state. The notation $z_{it} \in \{s_1, \ldots, s_J\}$ will represent the true state of individual i on occasion t, thus $N_t = \sum_{i=1}^{N} I(1 < z_{it} < J)$, where $I(\ldots)$ is and indicator function.

Schwarz and Arnason (1996) and Crosbie and Manly (1985) parameterized entry in the JS model using multinomial probabilities, ξ_t, $t = 1, \ldots, K$, which are the unconditional probabilities that an individual enters between occasions $t-1$ and t and is available for capture on occasion t. With the ξ_t entry probabilities we can form the *conditional* entry

probabilities that will be used in the JS model computations herein,

$$\tilde{\xi}_t = \frac{\xi_t}{\sum_{j=t}^{K} \xi_j},$$

which is the probability that an individual enters between $t-1$ and t given it has not entered prior to $t-1$ but will enter before K. Note that we only need to model the ξ_t up to a normalizing constant to calculate $\tilde{\xi}_t$ because that is the form of entry that is actually used in the likelihood we will work with here. Because we are considering multistate models as well, when an individual first enters the available population between $t-1$ and t we must also model which state it enters into. These are denoted, $\alpha_{l,t}$, where $\sum_{l=2}^{J-1} \alpha_{l,t} = 1$.

We turn our attention to transitions between states (including exit) after entry. Given that an individual has entered prior to occasion t, the probability that an individual does not exit the population between occasions t and $t+1$ given it is currently available for capture in state s_l on occasion t is $\phi_{l,t}$, $l = 2, \ldots, J-1$ and $t = 1, \ldots, K-1$. That is, the probability that an individual transitions from $s_l \rightarrow s_J$ is $1 - \phi_{l,t}$. If we consider s_J to be "death" then the ϕ parameters are the typical survival probabilities. Finally, $\psi_{l,l',t}$ ($\sum_{l'=2}^{J-1} \psi_{l,l',t} = 1$) is the probability that an individual transitions from state s_l to $s_{l'}$ between times t and $t+1$ given it was in the available population at time t and survived to occasion $t+1$.

Now that we have described how individuals transition in and out of availability and through states we can consider a description of the data models for the observed data. Traditionally, data for the JS model are composed of capture histories for each individual captured at least once, $\mathbf{x}_i = (x_{i1}, \ldots, x_{iK})$ where $x_{it} \in \{0, 1\}$ depending on whether or not individual $i = 1, \ldots, n$ was captured on occasion t. However, we expand this notation in a general way to cover other models such as robust design and multistate JS models. For example, x_{it} might represent the number of times the animal was captured in R_t secondary occasions during primary occasion t which defines a robust design model (e.g., Kendall et al. 1995). Or, x_{it} might be recorded as the individual being in a particular state if it is captured, implying a multistate model (White et al., 2006). We denote the likelihood of observing x_{it} given the individual is in state $z_{it} = s_l$ as $[x_{it}|s_l]$. Throughout the paper, we use the "$[A|B]$" notation to represent a conditional density (distribution) function of A given B. We briefly present a few examples for common models. For a standard Jolly-Seber model with a binary capture history and only 1 available state (s_2 = "alive") the likelihood would be,

$$[x_{it}|s_2] = \text{Bernoulli}(x_{it}|p_t),$$

where p_t is the probability of capture on occasion t. If the model being fit is a robust design formulation where x_{it}, are the number of captures in R_t sub-occasions within each

main occasion, the likelihood would be,

$$[x_{it}|s_2] = \text{Binomial}(x_{it}|R_t, p_t).$$

For a multistate model where x_{it} is recorded as a categorical state with uncertainty (e.g., Johnson et al. 2016) then we use the mixture distribution

$$[x_{it}|s_l] = (1 - p_{t,l})I(x_{it} = 0) + p_{t,l}\text{Categorical}(x_{it}|\boldsymbol{\delta}_l, x_{it} \neq 0)$$

where $\boldsymbol{\delta}_l$ is the vector of probabilities for the range of categories that can be observed when the individual is captured in state s_l. For example, one might accurately observe the state $x_{it} = s_l$ or it is unknown, say $x_{it} =$"u", then δ_l would represent the probability of observing the true state. For any model $[x_{it}|s_1(\text{or } s_J)]$ is an indicator that $x_{it} = 0$.

2.1 A brief review of data augmentation for Jolly-Seber models

To provide a framework for developing the collapsed Gibbs sampler that does not require data augmentation, we provide a brief review of the PX-DA approach for JS inference. Let $\mathbf{x}$ represent the $n \times K$ matrix of observed data, $\mathbf{z}$ is the associated matrix of true state values. $\mathbf{x}_{aug} = \mathbf{0}$ and $\mathbf{z}_{aug}$ are the $M - n \times K$ matrices for the $M - n$ augmented individuals. For the PX-DA approach we combine, $\mathbf{x}^+ = (\mathbf{x}', \mathbf{x}'_{aug})'$ and $\mathbf{z}^+ = (\mathbf{z}', \mathbf{z}'_{aug})'$, and treat them as observed data and true states respectively. The indicator of population inclusion is $\mathbf{w}^+ = (\mathbf{1}', \mathbf{w}'_{aug})'$. The first n components are known to be 1 because those individuals are observed.

The posterior distribution for inference is

$$[\boldsymbol{\theta}, \psi, \mathbf{w}^+, \mathbf{z}^+|\mathbf{x}^+] \propto [\mathbf{x}^+|\mathbf{z}^+, \mathbf{w}^+, \boldsymbol{\theta}][\mathbf{z}^+|\boldsymbol{\theta}][\mathbf{w}^+|\gamma][\boldsymbol{\theta}][\gamma], \tag{1}$$

where $\boldsymbol{\theta}$ are the parameters of the JS model described in the previous section and γ is the probability that an individual is a member of the superpopulation.

The main benefit of the PX-DA approach is that all the distributions on the right-hand side of (1) are all standard distributions, so they can readily be described in the model statements for MCMC software such as `JAGS` and `nimble` and the dimension of the products remain fixed at M individuals so trans-dimensional methods are not needed. Gibbs sampling for the PX-DA proceeds by sequentially sampling from the conditional distributions: $[\mathbf{z}^+|\boldsymbol{\theta}, \mathbf{w}^+, \mathbf{x}^+]$, $[\mathbf{w}^+|\boldsymbol{\theta}, \gamma, \mathbf{z}^+, \mathbf{x}^+]$, $[\gamma|\mathbf{w}^+]$, and $[\boldsymbol{\theta}|\mathbf{z}^+, \mathbf{w}^+, \mathbf{x}^+]$. Of course, these may be broken up further into scalar parameter updates. During updates, the raw quantities $\mathbf{z}^+$ and $\mathbf{w}^+$, as well as, γ are not usually useful on their own and abundance metric summaries, say $\boldsymbol{\vartheta} = f(\mathbf{z}^+, \mathbf{w}^+)$, are usually stored such as $N = \mathbf{w}'\mathbf{1}$ or

$N_t = \sum_{i=1}^{M} z_{it} w_i$. So, the resulting posterior desired is usually, $[\boldsymbol{\theta}, \boldsymbol{\vartheta}|\mathbf{x}^+]$. Note, that the posterior distribution used for inference depends on the augmentation used, so it is not the exact posterior we might want. If M is chosen large enough, however, it is a very good approximation.

3 Collapsed Gibbs sampling for Jolly-Seber models

In general collapsed Gibbs samplers are constructed by marginalizing over some components of the model, using a standard Gibbs sampler to sample the remaining parameters, then the marginalized parameters are sampled using their posterior predictive distribution given the first sample. In an MCMC sampler with strictly Gibbs updates, faster convergence is achieved because the first sample is not conditioned on the second sample components, reducing the covariance between them and speeding convergence (Liu, 1994). To develop the collapsed JS sampler we will marginalize over all the augmentation components so that we only sample from the conditional distribution,

$$[\boldsymbol{\theta}, \lambda|\mathbf{x}] \propto [\mathbf{x}|\boldsymbol{\theta}, \lambda][\boldsymbol{\theta}][\lambda],$$

where λ is a parameter that controls the overall superpopulation size. The origin of λ will be described in the following section, but it takes the place of γ when not using augmentation.

After drawing a sample from $[\boldsymbol{\theta}, \lambda|\mathbf{x}]$ we proceed by next drawing a sample from the posterior predictive distribution,

$$[\mathbf{z}, \mathbf{z}_u, n_u|\boldsymbol{\theta}, \lambda, \mathbf{x}]$$

where n_u is the posterior predicted number of unobserved individuals and $\mathbf{z}_u = [z_{u,jt}]$ is the associated $n_u \times K$ state matrix for those latent individuals. One important aspect to note is that the dimension of $\mathbf{z}_u$ is not constant over MCMC iterations. Therefore one usually saves the same fixed-dimensional abundance metrics as in the PX-DA scenario, $\boldsymbol{\vartheta} = f(\mathbf{z}, \mathbf{z}_u, n_u)$. Note the function, f is slightly different than before. For example, $N = n + n_u$ and $N_t = \sum_{i=1}^{n} z_{it} + \sum_{j=1}^{n_u} z_{u,jt}$. The resulting posterior of this sampler is

$$[\boldsymbol{\theta}, \boldsymbol{\vartheta}, \lambda|\mathbf{x}] \propto [\boldsymbol{\vartheta}|\boldsymbol{\theta}, \lambda, \mathbf{x}][\boldsymbol{\theta}, \lambda|\mathbf{x}].$$

Note that the posterior depends only on $\mathbf{x}$ not $\mathbf{x}^+$, so there is no approximation due to data augmentation.

Now that we have outlined the collapsed Gibbs sampler for JS models we need to examine the details and illustrate that it can be accomplished within the `nimble`

framework in way that makes it similar in ease-of-use for practitioners relative to the PX-DA approach. The first step is to determine the first sample marginal distribution.

3.1 Evaluating marginal posterior of Jolly-Seber models

In the first step we need to draw a sample from the marginal posterior distribution, $[\boldsymbol{\theta}, \lambda|\mathbf{x}]$, so let us start with the PX-DA posterior and marginalize over the augmented components. To accomplish this we follow King et al. (2016) and Hooten et al. (2023) in partitioning the PX-DA posterior by conditioning on capturing n individuals with capture-histories, $\mathbf{x}$,

$$\begin{aligned}[\boldsymbol{\theta}, \gamma, \mathbf{w}_{aug}, \mathbf{z}^{+}|\mathbf{x}] \propto [\mathbf{x}|\mathbf{z}, \boldsymbol{\theta}, n][\mathbf{z}|\boldsymbol{\theta}, n] \\ \times [\mathbf{z}_{aug}, \mathbf{w}_{aug}|\boldsymbol{\theta}, \gamma, n] \\ \times [n|\boldsymbol{\theta}, \gamma][\boldsymbol{\theta}][\gamma]\end{aligned}$$

where, under the PX-DA approach, the probability of observing n individuals during the study is $[n|\boldsymbol{\theta}, \gamma] = \text{Binomial}(M, \gamma p^*)$ and p^* is the probability of being captured at least once. Thus if we integrate over the augmentation components we obtain

$$\begin{aligned}[\boldsymbol{\theta}, \gamma, \mathbf{z}|\mathbf{x}] &\propto \sum_{\mathbf{w}_{aug}} \sum_{\mathbf{z}_{aug}} [\boldsymbol{\theta}, \gamma, \mathbf{w}_{aug}, \mathbf{z}^{+}|\mathbf{x}^{+}] \\ &\propto [\mathbf{x}|\mathbf{z}, \boldsymbol{\theta}, n][\mathbf{z}|\boldsymbol{\theta}, n][n|\boldsymbol{\theta}, \gamma][\boldsymbol{\theta}][\gamma]\end{aligned}$$

Before we continue with marginalizing over $\mathbf{z}$ let us take a closer look at $[n|\boldsymbol{\theta}, \gamma]$. We do not have any augmented data in the marginal posterior, but we still have the remnants of that process, M and γ. One way to remove their effect is to envision a hypothetical infinite amount of augmented data, that is $M \to \infty$. If we follow this limit while simultaneously allowing $\gamma \to 0$ such that $M\gamma \to \lambda$, then $[n|\boldsymbol{\theta}, \gamma] \to [n|\boldsymbol{\theta}, \lambda] = \text{Poisson}(\lambda p^*)$. So, fully marginalizing over the components of an infinitely large augmented data set, yields the posterior,

$$[\boldsymbol{\theta}, \lambda, \mathbf{z}|\mathbf{x}] \propto [\mathbf{x}, \mathbf{z}|\boldsymbol{\theta}, n][n|\boldsymbol{\theta}, \lambda][\boldsymbol{\theta}][\lambda].$$

This is equivalent in form to the likelihood derived by Glennie et al. (2019) and Zhang et al. (2023) for spatial capture-recapture models when individuals are distributed as a spatial Poisson process.

Now the final step in evaluating the fully marginal posterior distribution,

$$[\boldsymbol{\theta}, \lambda|\mathbf{x}] \propto [\mathbf{x}|\boldsymbol{\theta}, n][n|\boldsymbol{\theta}, \lambda][\boldsymbol{\theta}][\lambda]$$

is to marginalize over the partially latent states, $\mathbf{z}$. Glennie et al. (2019) and McClintock et al. (2020) show an individual capture history in a JS models can be formulated as a

Hidden Markov Model (HMM). That is, the (unconditional) probability of an individual capture-history can be calculated with the matrix product,

$$[\mathbf{x}_i|\boldsymbol{\theta}] = \boldsymbol{\pi}\mathbf{P}_1(x_{i1})\boldsymbol{\Gamma}_1\mathbf{P}_2(x_{i2})\cdots\boldsymbol{\Gamma}_{K-1}\mathbf{P}_K(x_{iK})\mathbf{1}. \tag{2}$$

where $\boldsymbol{\pi}$ is the initial state probability vector, $\mathbf{P}_t(x)$ is a diagonal matrix of observation likelihoods for each state, $\boldsymbol{\Gamma}_t$ is the probability transition matrix for states, and $\mathbf{1}$ is a vector of ones. Therefore, to evaluate the probability of a single capture-history, $\mathbf{x}_i$, conditional on being captured at least once, we can use the HMM structure and evaluate,

$$[\mathbf{x}_i|\boldsymbol{\theta}, n] = \frac{[\mathbf{x}_i|\boldsymbol{\theta}]}{1 - [\mathbf{x}_i = \mathbf{0}|\boldsymbol{\theta}]} \tag{3}$$

where the denominator, $p^* = 1 - [\mathbf{x}_i = \mathbf{0}|\boldsymbol{\theta}]$. We can use efficient HMM algorithms to evaluate both the numerator and denominator because the denominator is calculated by simply replacing the observed capture-history with all 0s and evaluating it with (2) as well.

For the general JS family model the matrices in the likelihood function (3) are given as follows. First, the initial state probability vector is

$$\boldsymbol{\pi} = \begin{array}{c} \begin{array}{cccccc} s_1 & s_2 & \cdots & s_{J-1} & s_J \end{array} \\ \begin{bmatrix} 1-\tilde{\xi}_1 & \tilde{\xi}_1\alpha_{1,2} & \cdots & \tilde{\xi}_1\alpha_{1,J-1} & 0 \end{bmatrix} \end{array}.$$

The observation likelihood matrix, $\mathbf{P}_t(x_{it})$ is a $J \times J$ diagonal matrix with entries, $[x_{it}|s_l]$, $l = 1, \ldots J$. Finally, the probability transition matrix is, for $t = 1, \ldots, K-1$,

$$\boldsymbol{\Gamma}_t = \begin{array}{c} \begin{array}{cccccc} s_1 & s_2 & \cdots & s_{J-1} & s_J \end{array} \\ \begin{bmatrix} 1-\tilde{\xi}_{t+1} & \tilde{\xi}_{t+1}\alpha_{t+1,2} & \cdots & \tilde{\xi}_{t+1}\alpha_{t+1,J-1} & 0 \\ 0 & \phi_{t,2}\psi_{t,2,2} & \cdots & \phi_{t,2}\psi_{t,2,J-1} & 1-\phi_{t,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \phi_{t,J-1}\psi_{t,J-1,2} & \cdots & \phi_{t,J-1}\psi_{t,J-1,J-1} & 1-\phi_{t,J-1} \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \begin{array}{c} s_1 \\ s_2 \\ \vdots \\ s_{J-1} \\ s_J \end{array} \end{array}. \tag{4}$$

Readers should note that we have left the portion of the original JS models dealing with death upon capture out of this development. We find this part is often omitted in practice. If needed, it can be included in the model using a second exit state in the transition matrix.

3.2 Using `nimble` to perform collapsed Gibbs sampling

To perform this approach in `nimble` we start by forming custom distributions for the model description. Unlike `jags`, `nimble` possesses the ability to define new distributions

and other functions that can be directly incorporated in model statements for MCMC analysis. This capability is accessed through `nimble::nimbleFunction()`. For categorical data (e.g., multistate capture-histories), the `nimbleEcology` package has already defined an unconditional HMM distribution (2). This is accessed via the `nimbleEcology::dDHMMo()` HMM distribution. In the `nimbleJSextras` package we used this general HMM distribution to create the bespoke JS distribution functions,

```
dJS_ms(x[i,1:K],...,pstar,) = dDHMMo(x[i,1:K],...) / pstar
```

where `pstar = 1 - dDHMMo(zeros[1:K],...)` and `zeros` is a fixed vector of zeros. The arguments in "..." are the $\boldsymbol{\pi}$, $\mathbf{P}_t(x_{it})$ and $\boldsymbol{\Gamma}_t$ matrices as specified by the `nimbleEcology::dDHMMo()` function. The the overall probability of capture, `pstar`, has been made a separate argument because it is the same for all individuals and it is unnecessary to recalculate the same quantity for each observed individual. The `nimbleJSextras` package also has a `pstar_ms()` function to calculate this so the user does not have to unnecessarily create the `zeros` vector.

Now that the `dJS_ms()` distribution and the `pstar_ms()` function are defined, we can it can be used in a `nimble` model statement along with the Poisson distribution to model the number of captured individuals, for example,

```
for(i in 1:nobs) {
  x[i,1:K] ~ dJS_ms(...,pstar)
}
pstar <- pstar_ms(...)
n ~ dpois(lambda * pstar)
```

where `x` is a $n \times K$ matrix for which x_{it} = `x[i,1:K]`. Here, `nobs` $= n$, but we can not pass n as both a constant and a random variable, so we just use a different name for the i index limit. Outside of specifying the HMM matrices for particular JS models, this code is all that is necessary to estimate the model parameters with MCMC.

Now we turn our focus to sampling the abundance functions $\boldsymbol{\vartheta} = f(\mathbf{z}, \mathbf{z}_u, n_u)$ in the second stage of the collapsed Gibbs sampler by drawing from

$$[\mathbf{z}, \mathbf{z}_u, n_u | \boldsymbol{\theta}, \lambda, \mathbf{x}] = [n_u | \boldsymbol{\theta}, \lambda, \mathbf{x}][\mathbf{z}_u | n_u, \boldsymbol{\theta}, \mathbf{x}][\mathbf{z} | \boldsymbol{\theta}, \mathbf{x}]. \tag{5}$$

To begin, because $[N|\lambda] = \text{Poisson}(\lambda)$ under the same infinite augmentation scheme and $[n, n_u | N, \boldsymbol{\theta}] = \text{Multinomial}(N, p^*, 1 - p^*)$ implies

$$[n_u | \boldsymbol{\theta}, \lambda] = \text{Poisson}(\lambda(1 - p^*))$$

conditionally independent of the data, $\mathbf{x}$. So, for each iteration of an MCMC algorithm it can be updated with a draw from a standard Poisson distribution as a posterior predictive variable. Then the prediction for the overall abundance, N, results directly from $N = n_u + n$. Schofield et al. (2023) and Johnson et al. (2010) both describe abundance estimation as *predicting* the number of undetected individuals. If one is only interested in the demographic rate parameters, $\boldsymbol{\theta}$, N, or derived quantities of these parameters, then one only needs to add

```
nu ~ dpois(lambda*(1-pstar))
Nsuper <- nobs + nu
```

to the previous `nimble` model statement. If quantities such as N_t or population growth are desired, we must sample $\mathbf{z}$ and $\mathbf{z}_u$ to derive these within the sampler.

The final step in the collapsed sampler is to sample the unobserved and partially observed state matrices, $\mathbf{z}_u$ and $\mathbf{z}$. Because we have marginalized over the true states, $\mathbf{z}$ with the HMM-based likelihood calculations it is not as readily apparent how this can be accomplished, however, we can use standard HMM algorithms to efficiently draw from those posterior distributions as well. Then we can extend the method to $\mathbf{z}_u$.

To begin the final step, we must sample the true states for the observed individuals from the predictive distribution, $[\mathbf{z}_i|\boldsymbol{\theta}, \mathbf{x}_i]$. This can be accomplished efficiently with the Forward–Filtering Backward–Sampling (FFBS) algorithm for conditional state sampling of HMMs (Scott, 2002). The details of the of the FFBS algorithm can be found in Appendix A (algorithm 1). Heuristically, the algorithm works by first using the HMM forward algorithm (Zucchini et al., 2016) followed by a backward pass through the capture history to sample from $[z_{it}|z_{i,t+1}, \ldots, z_{i,K}, \boldsymbol{\theta}, \mathbf{x}_i]$ at each occasion. To facilitate this sampling algorithm we have coded the FFBS algorithm into the `nimbleJSextras` package and it can be implemented via,

```
for(i in 1:nobs){
  z[i,1:K] <- sample_det_ms(x=x[i,1:K], ...),
  }
```

this will draw a sample from $[\mathbf{z}_i|\boldsymbol{\theta}, \mathbf{x}_i]$. Here we have proceeded against convention by specifying a random quantity with a deterministic `nimble` function, however, one would never use this distribution as a likelihood, so the coding overhead to specify it as a distribution seems unnecessary and it functions in the same manner as a bespoke distribution. Once we perform the the FFBS algorithm for each captured individual we can calculate the derived quantity, $n_{t,l} = \sum_{i=1}^{n} I(z_{it} = s_l)$, the abundance of observed

individuals in state s_l at time t. For simplicity, we use the bold notation, $\mathbf{n}_t = (n_{t,1}, \ldots, n_{t,J})$, to represent the vector of state-specific abundances and $n_t = \sum_{l=2}^{J-1} n_{t,l}$ to represent the total abundance of captured individuals available at time t. For a standard JS model ($J = 3$) this is accomplished with the following code.

```
available[1:nobs,1:K] <- z[1:nobs,1:K]==2
for(t in 1:K){ n_t[t] <- sum(available[1:nobs, t]) }
```

Now that we have the state and time-specific abundances for the observed individuals, we shift to the n_u individuals that are never captured. We could use the same FFBS algorithm to individually sample the states of all n_u individuals all with the capture-history $\mathbf{x}_i = \mathbf{0}$. However, Because all uncaptured individuals have the same capture-history, we can use a multinomial version of the FFBS, to sample the state and time-specific abundance in one forward and backward pass without simulating individual state histories (Appendix A, algorithm 2). This can be accomplished using `nimbleJSextras` with the model statement

```
nu_t[1:3, 1:K] <- sample_undet_ms(nu,...)
```

Now that we have posterior samples of $\mathbf{n}_t$ and $\mathbf{n}_{u,t}$ we can predict the state and time-specific abundances $\mathbf{N}_t = \mathbf{n}_t + \mathbf{n}_{u,t}$ and $N_t = n_t + n_{u,t}$ and any function, $f(\mathbf{N}_1, \ldots, \mathbf{N}_K)$, such as population growth. The online supplement provides a `nimble` code template for fitting a standard JS model via the collapsed sampler. Notably, the sample code illustrates the structure for the data input and `nimbleEcology::dDHHMo()` HMM matrices contained in the "`...`" arguments of the previous code snippets.

3.3 Additional computation benefits of the collapsed sampler

Here we demonstrate some additional computational savings that we can use due to the factoring of the posterior distribution in the collapsed Gibbs sampler. Using the previous development, we can write the full JS model posterior distribution as,

$$[\boldsymbol{\vartheta}, n_u, \boldsymbol{\theta}, \lambda, |\mathbf{x}] \propto [\boldsymbol{\vartheta}|n_u, \boldsymbol{\theta}, \mathbf{x}][n_u|\boldsymbol{\theta}, \lambda, \mathbf{x}] \times [\mathbf{x}|\boldsymbol{\theta}, n][n|\boldsymbol{\theta}, \lambda][\boldsymbol{\theta}][\lambda] \tag{6}$$

The first savings to note is that we only need to use the unique capture histories to evaluate the fully marginal likelihood (Turek et al., 2016),

$$[\mathbf{x}|\boldsymbol{\theta}, n] = \prod_{i=1}^{n} [\mathbf{x}_i|\boldsymbol{\theta}, n] = \prod_{j=1}^{n^*} w_j [\mathbf{u}_j|\boldsymbol{\theta}, n] \tag{7}$$

where $\mathbf{u}_j$ is the jth unique capture-history, w_j is the number of times it occurs in $\mathbf{x}$ and n^* is number of unique histories in $\mathbf{x}$. This is standard practice when fitting models using

maximum likelihood methods, however, it is not possible using the PX-DA approach for Bayesian analysis. This can be a substantial computational savings if n^* is much smaller than n (Turek et al., 2016).

The second computational saving is due to the predictive factoring in (6). This is the same approach used by Hooten et al. (2023), and Hooten et al. (2024) to facilitate an efficient 2-stage approach to Bayesian estimation of closed-population models and we can do the same here. The 2-stage approach first involves drawing a posterior sample from $[\boldsymbol{\theta}, \lambda|\mathbf{x}, n]$, preferably using just the unique capture-histories. Then, divide the sample into B batches and sample from $[\boldsymbol{\vartheta}, n_u|\boldsymbol{\theta}, \lambda, \mathbf{x}]$ in parallel for each $(\boldsymbol{\theta}, \lambda)$ in the separate batches.

4 Examples and Extensions

4.1 The Dipper Data

The first examination of the ubiquitous European dipper data was presented by Cormack (1964) who used it for demonstration of the Cormack-Jolly-Seber (CJS) model for survival estimation from capture-recapture data. Many years later Lebreton et al. (1992) used the data to demonstrate a JS analysis for abundance estimation. More recently, Royle and Dorazio (2008) demonstrate a Bayesian version of JS modeling using PX-DA. Here we demonstrate analysis of the data using the *per capita* recruitment parameterization of Link and Barker (2005) for modeling entry probability.

4.1.1 Model specifications

The entry probabilities of the *per capita* recruitment model are given by

$$\xi_t \propto \begin{cases} 1 & t = 1 \\ d_t f_{t-1} & t = 2, \ldots, K \end{cases}$$

$d_t \propto E[N_t/N|\boldsymbol{\theta}, N]$, the expected relative abundance, and f_t is the *per capita* recruitment rate. The quantity d_t is calculated with the recursion, $d_t = d_{t-1}(\phi_{t-1} + f_{t-1})$ where $d_1 \equiv 1$.

In addition to the *per capita* recruitment parameterization of the entry model, we also use a hierarchical approach for estimation of the parameters. The hierarchical model adds the random effects components,

$$\text{logit } \phi_t \sim N(\mu_\phi, \sigma^2_\phi)$$
$$\text{logit } p_t \sim N(\mu_p, \sigma^2_p), \text{ and } \log f_t \sim N(\mu_f, \sigma^2_f)$$

This allows shrinkage of the full time model to the temporally constant model. We used the complexity prior distribution (Simpson et al., 2017) on each of the three the variance components, $\sigma \sim \text{Exponential}(1)$. All mean components of the random effect where given the vaguely informative $N(0, 1.5^2)$ prior distribution. For λ we used $[\lambda] = \text{Gamma}(1.0\text{e-}6, 1.0\text{e-}6)$ to approximate the objective prior, $[\lambda] = \lambda^{-1}$(Schofield et al., 2023).

Code for this example is provided in the online supplement and the dipper data was obtained from the `marked` package (Laake et al., 2013). We analyzed this same model with four different methods: (1) PX-DA with 300 augmented individuals, (2) PX-DA with 1,000 augmented individuals, (3) the collapsed sampler, and (4) the collapsed sampler using additional computational savings of Section (3.3). For the 2-stage collapsed sampler, we used 10 parallel sessions so that each session used 5,000 iterations from the first stage to sample abundance. We compared the first 5,000 iterations of methods to assess progress toward convergence using the scale reduction diagnostic (Gelman and Rubin, 1992) as well as the multivariate version (Brooks and Gelman, 1998). For full estimation we used 50,000 MCMC iterations following a 10,000 iteration burnin.

4.1.2 Results

The multivariate scale reduction for the first 5,000 iterations of the collapsed approach was 1.11, while the scale reduction for the PX-DA approach with 300 and 1,000 augmented individuals was 1.22 and 1.33 respectively. For this diagnostic a value ≈ 1 implies convergence. So, we can see that convergence begins sooner for the collapsed sampler, but for this analysis it was of little consequence because all methods had converged well before the 10,000 iteration burnin.

To compare the relative run-times of each of the methods we hold PX-DA (300) as the default standard. The extra 700 augmented individuals in the second PX-DA increased run-time by 3.68×. Run-time for the standard collapsed sampler was 1.82×. Finally, the run-time of the two-stage parallel collapsed sampler was 0.11×, a substantial improvement over the PX-DA methods. For the dipper data there are only $n^* = 32$ unique capture-histories in the $n = 294$ observed individuals. So, the likelihood calculation only needs to loop over the 32 unique histories for each MCMC iteration.

Parameter estimates are given in Table 1 for the standard collapsed sampler. For each of the four methods, the parameter and abundance estimates aligned outside of Monte Carlo error, as expected, because the same model is fitted. The super population size estimate is $N = 324.6$ with SD = 16.3 and 95% highest probability density interval of [300–356]. Yearly abundance estimates are illustrated in Figure 1.

4.2 Abundance and reproductive demographics of nesting Hawaiian green sea turtles

Here we present an example analysis with a multistate component, abundance of female nesting Hawaiian green sea turtles (*Chelonia mydas*) or *honu*. This analysis demonstrates use of the fully marginal collapsed sampler to fit a model that incorporates individual by time effects on capture probability heterogenity. Specifically we fitted a behavioral model with effect of previous capture on the current capture probability. In addition, these data are relatively large and PX-DA methods can be computationally expensive in that case. So, we can use the two-stage collapsed sampler to efficiently sample the posterior.

The data presented here are from the 1973–2016 nesting season surveys on the islets within the atoll of Lalo (French Frigate Shoals), in the Northwestern Hawaiian Islands. Over the 44 year study period n = 5,271 females were captured. During the season, honu will migrate from the Main Hawaiian Islands (MHI) to nest at Lalo starting in late April through early August. After nesting is complete honu will return to the MHI and enter the foraging phase which lasts between 2–10 (average *c.* 4) years, where no nesting occurs (Balazs et al., 2015). The time between nesting occasions is termed *remigration interval*.

4.2.1 Model Specifications

There are two alterations to the traditional JS model that are necessary to estimate the size of the nesting female population with this data. The first issue is that individuals become unavailable for capture for an extended time while they are foraging as it is not feasible to capture individuals in the MHI with any systematic effort. During nesting phase individuals are concentrated at Lalo, thus, are relatively easy to capture. Second, processing individuals on their initial capture (i.e., unmarked animals) is more time consuming than recording individuals that already have a permanent identification (marked). In some years there is not enough total survey resources to equally process marked and unmarked individuals. Therefore, in some years capture and processing of unmarked individuals is lower priority than recording marked individuals. This is similar to a closed-population M_b model (Otis et al., 1978) where the individuals become "trap happy", however, this mathematical behavioral response is due to capture effort on the part of the researchers, not actual honu behavior.

To model the effect of the remigration interval on abundance of hawksbill marine turtles (*Eretmochelys imbricata*) Kendall et al. (2019) used the breeding return time (JS-BRT) model of Pledger et al. (2013) which is a special case of the semi-Markov capture-recpature model described by King and Langrock (2016). The JS-BRT model uses

several intermediate (sub-) states to represent the overall condition of the individual being in a nonbreeding state, e.g., for honu, the individual is in a foraging (nonbreeding) state if it is in any one of the substates, $F_1, \ldots,$ or F_m. The transition probabilities force the individual to pass through the substates sequentially or return to the reproductive (nesting) state. With regard to the composite state, $F = \{F_1, \ldots, F_m\}$, the underlying model is a Hidden *semi*-Markov model (HSMM), because the dwell-time in the composite state is not geometrically distributed as it would be in a Markov chain.

Now we examine the effort-induced trap response issue. In a PX-DA analysis this would be taken care of with a individual $\times$ occasion covariate, say c_{it} that indicates if individual i has been previously captured on occasion t. We cannot use that approach in the collapsed sampler because we are not explicitly simulating state-histories for uncaptured individuals. However, we can use the another set of substates to define our reproductive state: R_{00}, R_{01}, R_{10}, R_{11}, where the subscripts indicate, respectively, previously captured and currently captured on occasion t. We have moved the capture process into the state transitions. In addition to the composite reproductive state, we also need to further expand the foraging substates to $F_{01}, \ldots, F_{0m}, F_{11}, \ldots, F_{1m}$ where the first subscript denotes previous capture. These additional foraging states are not biologically different, they simply serve to prevent transitions among the reproductive states that are not permitted, e.g., if an individual is in R_{00} its next reproductive state cannot be R_{11} or R_{10}.

The model specifications for this model are notably different from the standard JS form described previously. However, It can still be placed in the same HMM framework for which we can use the collapsed Gibbs sampler with the additional computational savings benefits. The HMM matrix specifications are as follows.

We begin with the state transition matrix which can be represented in its general form with the block structure,

$$\mathbf{\Gamma}_t = \begin{bmatrix} 1-\tilde{\xi}_{t+1} & \tilde{\xi}_{t+1}\mathbf{\Psi}_{\emptyset,R,t} & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{0} & \phi_{R,t}\mathbf{\Psi}_{R,F,t} & (1-\phi_{R,t})\mathbf{1} \\ \mathbf{0} & \phi_{F,t}\mathbf{\Psi}_{F,R,t} & \phi_{F,t}\mathbf{\Psi}_{F,F,t} & (1-\phi_{F,t})\mathbf{1} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

where the $\mathbf{\Psi}_{a,b,t}$ matrices represent the submatrix describing transitions from the composite state "a" to composite state "b" (Langrock and Zucchini, 2011; King and Langrock, 2016). The entry 1×4 entry submatrix is, $\mathbf{\Psi}_{\emptyset,R,t} = [1-\eta_{t+1}p_{t+1} \quad \alpha_{t+1}p_{t+1} \quad 0 \quad 0]$, where p_t is the capture-probability and $0 \leq \eta_t \leq 1$ is the proportional reduction in capture probability due to reduced effort for capturing unmarked individuals. The submatrix $\mathbf{\Psi}_{R,F,t}$ is a $4 \times 2m$

indicator matrix such that R_{00} transitions to F_{01} and R_{01}, R_{10}, and R_{11} transition to F_{11} automatically because honu cannot nest in subsequent years. The $\mathbf{\Psi}_{F,R,t}$, matrix is composed of elements,

$$\psi_{F_{0j},R_{00},t} = \tau_j(1-\eta_{t+1}p_{t+1}), \qquad \psi_{F_{0j},R_{01},t} = \tau_j\eta_{t+1}p_{t+1}$$

$$\psi_{F_{1j},R_{10},t} = \tau_j(1-p_{t+1}), \qquad \psi_{F_{1j},R_{11},t} = \tau_j p_{t+1}$$

for $j = 1, \ldots, m$ where τ_j is the probability of transitioning from foraging substate j to a reproductive state. All other F to R transitions have probability zero. Finally, the $\mathbf{\Psi}_{F,F,t}$ matrix is composed of elements,

$$\psi_{F_{0j},F_{0,j+1},t} = 1-\tau_j \text{ for } j = 1, \ldots, m-1, \text{ and } \psi_{F_{0m},F_{0m},t} = 1-\tau_m.$$

The F_1 transitions are identical and all other probabilities are 0.

The bulk of the model specification falls in the transition matrix. The other HMM matrices are, comparatively, more straightforward. The initial entry vector is,

$$\boldsymbol{\pi} = \begin{bmatrix} 1-\tilde{\xi}_1 & \tilde{\xi}_1(1-\eta_1) & \tilde{\xi}_1\eta_1 & \mathbf{0} \end{bmatrix}$$

where $\mathbf{0}$ is a $2m+3$ row vector of zeros. The observation matrices, $\mathbf{P}(x_{it})$, are still diagonal with $[x_{it}|s_l]$ entries. However, because the detection process has been absorbed in the states, $[x_{it}|s_l]$ is just an indicator that the observation x_{it} is possible in state s_l. For example $[x_{it}=0|s_l=R_{00}]=1$ and $[x_{it}=0|s_l=R_{01}]=0$.

Now that we have the general form of the HMM matrices, we can define specific components for this analysis. Honu are long-lived with very stable survival once they reach adulthood (Piacenza et al., 2016) so we modeled survival probability as constant over time and between states, $\phi_{l,t} = \phi$ with prior distribution $[\text{logit } \phi] = t(\sigma = 1.52, \text{df} = 6)$, where $t(\sigma, \text{df})$ is a scaled t distribution with df degrees of freedom. We used this specification to closely approximate a uniform distribution for ϕ but the sampled parameter, logit ϕ, is unconstrained. The entrance probabilities are modeled using recruitment probability (Royle and Dorazio, 2008). The recruitment probability, ρ_t, is the probability that the individual enters at occasion t given it has not yet entered. This essentially mirrors the exit probability. To use it in the collapsed sampler we need the unconditional entry probability,

$$\xi_t \propto \begin{cases} \rho_1 & t = 1 \\ \rho_t \prod_{j=1}^{t-1}(1-\rho_j) & t \geq 2 \end{cases},$$

from which we obtain $\tilde{\xi}_t$ for the transition matrices. We, again, used the logit transformation to specify the prior distributions $[\text{logit } \rho_t] = t(1.52, 6)$. For the capture

probability parameters, p_t and η_t we used the $t(1.52, 6)$ prior distributions as well. Finally, we used $m = 5$ foraging substates to model the remigration interval, with $[\text{logit } \tau_j] = t(1.52, 6)$. This choice of m results in a total of $J = 1 + 4 + 2(5) + 1 = 16$ states for the model. Finally, we used the approximate objective prior $[\lambda] = \text{Gamma}(1.0\text{e-}6, 1.0\text{e-}6)$.

Using the two-stage approach we ran the MCMC sampler for 50,000 iterations following a 20,000 iteration burn-in. Within the histories of the 5,271 observed individuals, there were only $1,338$ unique capture histories. So, we reduced the sample size burden in the first stage by about 75%. Anecdotally, as all computation environments are different, on an Apple MacBook Pro with an M2 Max processor, the first-stage MCMC took approximately 3.4 hours to run. For the second stage we used 10 parallel sessions and this completed in approximately 7.5 minutes. If we assume that there is linear run time relative to n, this implies the equivalent PX-DA version would take approximately 32 hours to complete. This is, of course, just an educated guess but illustrative of the possible improvement based on the dipper example.

4.2.2 Results

As expected, there were several years where capture probability was significantly lower for newly encountered individuals than those that had been previously captured (Figure 2). On some occasions $\hat{p}_t \approx 1$ while $\hat{\eta}_t \hat{p}_t$ was substantially lower. Yearly survival probability was $\hat{\phi} = 0.91$ [95% CI: 0.91 – 0.92]. The probability distribution function for remigration interval, r, can be calculated from the breeding return probabilities, τ_j using,

$$[r = k|\tau_1, \ldots, \tau_5] = \begin{cases} \tau_1 & k = 2 \\ \tau_{k-1} \prod_{j=1}^{k-1} (1 - \tau_j) & k \geq 3 \end{cases}$$

where for any index > 5, τ_5 is used. Remigration probability estimates peaked at 3 years and showed a slight decline to 4 years, illustrating the most common remigration interval is somewhere between 3 and 4 years (Figure 3). The average remigration from these probabilities is 3.74 years [3.68 – 3.82] which is similar to the empirically observed value of 4 reported by Balazs et al. (2015).

The estimates of recruitment probability (ρ_t) were low but relatively consistent for the bulk of the time series (Figure 4A) with a marked increase in both point estimates and uncertainty as the end of the series is reached. This is expected and is not indicative of a real biological increase in population entry, as can be seen by examining the corresponding estimates of marginal entry probability (ξ_t; Figure 4B). It is a result of the fact that the model constrains entry to occur before the end of the study, and the uncertainty results

from the breeding return time portion of the model. Individuals may enter near the end of the study and, if not captured, they most likely will not be seen before the study concludes.

Yearly abundance of female reproductive age honu and 90% credible intervals are shown in Figure 5. The first 6 years were discarded because there are females that have nested before and are in the desired population, but they were in the foraging phase and did not have a chance to be exposed to capture (Kendall et al., 2019). As Figure (5) illustrates, the population climbs rapidly for the first few years as all the foraging females become exposed to capture. The 6-year burn-in was chosen based on the posterior renesting interval distribution. For all MCMC samples, $< 1\%$ of renesting intervals last 7 years or more. So for those individuals that nested immediately before the first occasion, there is only a small chance that they did not nest again by the 7th occasion; thus, all females have been exposed to capture and N_t represents the abundance of all nesting females. The overall, superpopulation abundance was estimated $\hat{N} = 10{,}280$ [9,041–11,437]. The 99%-tile of the posterior was 12,217. Thus in order to make sure that M is large enough for a PX-DA version analysis, one should add at least 12,217 - 5,271 $\approx$ 7,000 individuals.

5 Discussion

Here we presented motivation and examples demonstrating that Bayesian inference for estimating abundance from capture-recapture data using Jolly-Seber family models does not necessarily need the PX-DA approach to be practical for regular use. By collapsing the PX-DA Gibbs sampler we can eliminate all need for state and data augmentation. Although there are no MCMC theories we can directly use to guarantee convergence improvements due to the Metropolis-within-Gibbs samplers that are used in practice. It is often noted that convergence improvements are realized in practice (Van Dyk and Jiao, 2015). In fact, in the dipper example we did see some initial convergence improvement, although only slight. However, more importantly, the collapsed sampler allows us to use some additional computational tricks which can drastically speed up the MCMC, only using unique capture histories and parallel sampling of any desired abundance quantities.

To make the collapsed approach straightforward for practitioners to use we developed the `R` package `nimbleJSextras`, which can be installed from Github or R-universe [†]. Both source and binary installations are available from R-universe. The package extends the `nimble` and `nimbleEcology` packages for fitting JS models using the collapsed approach. With `nimbleJSextras` fitting a JS models requires relatively the same level of bespoke

[†] `https://dsjohnson.r-universe.dev/nimbleJSextras`

coding as the PX-DA approach. In addition to the multistate models shown in the example section we have also included functions for fitting Binomial and Poisson capture histories as well. The code in these extra versions serves as a template for those that might want to extend to other types of capture-histories.

The examples of Section 4 illustrate both a standard JS analysis with the dipper data, as well as, inclusion of individual $\times$ occasion effects on capture heterogeneity in the honu analysis. But it is useful to consider when the PX-DA approach might be preferred. One of the main benefits of the PX-DA approach over even standard frequentist methods such as maximum likelihood estimation is the ease of which all types of heterogeneity can be included in a model. For example, individual heterogeneity in survival or capture probability can easily be included with an individual random effect within logit specifications of the parameters (see Royle and Dorazio 2008, Chapter 10). These continuously valued random effects would be difficult if not prohibitively challenging to implement with the collapsed approach. However, we speculate that, as in the honu example, there may be finite state alternatives or approximations for these models which could produce similar results. Pledger et al. (2010) note that finite state models with 2 states can be effective approximations for continuous capture heterogeneity models. However, a model with many different continuous random effects components might greatly expand the number of states necessary to approximate said model. So, while a single random effect for individual capture heterogeneity might be easily approximated by a few states, the number of states needed for heterogeneity in capture, survival, and state transitions will grow multiplicatively. In that case, PX-DA models may be easier to implement, but, some further investigation is warranted. For example, Johnson et al. (2016), show that multivariate states can be represented more parsimoniously than a full product of levels.

Astute readers will notice, while we used the collapsed Gibbs sampler to derive a fully marginal approach, in the end we arrived at a standard Bayesian analysis in which we use the exact data likelihood and prior distributions to arrive at posterior inference. For the true model parameters: $\mathbf{p}$, $\boldsymbol{\phi}$, $\boldsymbol{\xi}$, and λ, this is, indeed, the case. However, using the derivation of the collapsed sampler to arrive at this realization serves three purposes. First, it allows us to use some theoretical justification to provide some assurance that there is a very high likelihood that we will improve the MCMC mixing and convergence rate versus the PX-DA version, even if only slightly as in the dipper example. Yackulic et al. (2020) also demonstrated this for marginalized Cormack-Jolly-Seber models through simulation. Although, we are not mathematically guaranteed that this will occur. Second, the collapsed Gibbs sampler allows us to explicitly formulate abundance values, e.g.,

$N_1, \ldots, N_K$, as posterior predictive quantities, which is not immediately obvious from classical likelihood derivations such as Schwarz and Arnason (1996). This allows the use of the parallel sampling in the second stage. Finally, the process of deriving the collapsed Gibbs sampler allows us to connect the traditional frequentist JS likelihood to the the Bayesian PX-DA version, which for many practitioners has become the default way to think about abundance estimation from capture-recapture data. Given that we are using the exact likelihood in a traditional Bayesian analysis we can also examine some implications of our development in that light. The first thing to note is that N is not a parameter in our likelihood, λ takes its place. In the Schwarz and Arnason (1996) JS likelihood, $[n|\boldsymbol{\theta}, N] = \text{Binomial}(N, p^*)$ is used instead of $\text{Poisson}(\lambda p^*)$ (King et al., 2016). In a traditional Bayesian analysis using the likelihood we would put a prior distribution on N. If we assume $[N] = \text{Poisson}(\lambda)$ we will arrive at the same likelihood here, developed from allowing $M \to \infty$, after marginalizing over N. However, Link (2013) suggests the objective prior $[N] = N^{-1}$. But as Schofield et al. (2023) proves, using the Poisson prior with hyper-prior $[\lambda] = \lambda^{-1}$ will give the exact same posterior inference as when using the objective prior for N directly. So, while it may seem that we are making a specific distributional assumption for N by using the Poisson distribution, it is not a consequential as it seems. In the example sections we used $[\lambda] = \text{Gamma}(\epsilon, \epsilon)$ which approximates λ^{-1} for $\epsilon \to 0$. So, the collapsed development provides a framework to complete the arc of Bayesian inference for JS models from the exact likelihood of Schwarz and Arnason (1996) to the PX-DA expansion of Durban and Elston (2005) and Royle et al. (2007) and back again to use of the exact likelihood.

Acknowledgments

The findings and conclusions in the paper are those of the authors and do not necessarily represent the views of the National Marine Fisheries Service, NOAA. Reference to trade names does not imply endorsement by the National Marine Fisheries Service, NOAA. This manuscript has undergone National Marine Fisheries Service internal review for technical and policy content. Thanks to P.Conn for initial technical review and suggestions.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contributions

DSJ, SKI, and JJG developed proposed methodology; DSJ developed software for the collapsed sampler; DSJ and SKI performed data curation and formal analysis; DSJ led the writing of the original draft; SKI and JJG contributed to the review and editing of the manuscript.

References

Balazs, G. H., Van Houtan, K. S., Hargrove, S. A., Brunson, S. M., and Murakawa, S. K. (2015). A review of the demographic features of Hawaiian green turtles (*Chelonia mydas*). *Chelonian Conservation and Biology*, 14(2):119–129.

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455.

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

Crosbie, SF. and Manly, BFJ. (1985). Parsimonious modelling of capture-mark-recapture studies. *Biometrics*, 41(2):385–398.

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., and Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26(2):403–413.

Durban, JW. and Elston, DA. (2005). Mark-recapture with occasion and individual effects: Abundance estimation through Bayesian model selection in a fixed dimensional parameter space. *Journal of Agricultural Biological and Environmental Statistics*, 10(3):291–305.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.

Glennie, R., Borchers, D. L., Murchie, M., Harmsen, BJ., and Foster, RJ. (2019). Open population maximum likelihood spatial capture-recapture. *Biometrics*, 75:1345–1355.

Goldstein, B., Turek, D., Ponisio, L., and de Valpine, P. (2024). nimbleEcology: Distributions for ecological models in nimble. R package version 0.5.0, https://cran.r-project.org/package=nimbleEcology.

Hooten, M. B., Schwob, M. R., Johnson, D. S., and Ivan, J. S. (2023). Multistage hierarchical capture–recapture models. *Environmetrics*, 34:e2799.

Hooten, M. B., Schwob, M. R., Johnson, D. S., and Ivan, J. S. (2024). Geostatistical capture-recapture models. *Spatial Statistics*, 59:100817.

Johnson, D. S., Laake, J. L., Melin, S. R., and DeLong, R. L. (2016). Multivariate state hidden markov models for mark-recapture data. *Statistical Science*, 31:233–244.

Johnson, D. S., Laake, J. L., and Ver Hoef, J. M. (2010). A model-based approach for making ecological inference from distance sampling data. *Biometrics*, 66(1):310–318.

Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, 52(1/2):pp. 225–247.

Kendall, W. L., Pollock, K. H., and Brownie, C. (1995). A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. *Biometrics*, pages 293–308.

Kendall, W. L., Stapleton, S., White, G. C., Richardson, J. I., Pearson, K. N., and Mason, P. (2019). A multistate open robust design: Population dynamics, reproductive effort, and phenology of sea turtles from tagging data. *Ecological Monographs*, 89(1):e01329.

King, R. and Langrock, R. (2016). Semi-Markov Arnason-Schwarz models. *Biometrics*, 72:619–628.

King, R., McClintock, B. T., Kidney, D., and Borchers, D. (2016). Capture-recapture abundance estimation using a semi-complete data likelihood approach. *Annals of Applied Statistics*, 10(1):94–117.

Laake, J. L., Johnson, D. S., and Conn, P. B. (2013). Marked: An R package for maximum likelihood and Markov Chain Monte Carlo analysis of capture–recapture data. *Methods In Ecology and Evolution*, 4:885–890.

Langrock, R. and Zucchini, W. (2011). Hidden Markov models with arbitrary state dwell-time distributions. *Computational Statistics & Data Analysis*, 55(1):715–724.

669 Lebreton, J., Burnham, K. P., Clobert, J., and Anderson, D. R. (1992). Modeling survival
670 and testing biological hypotheses using marked animals: A unified approach with case
671 studies. *Ecological Monographs*, 62:67–118.

672 Link, W. A. (2013). A cautionary note on the discrete uniform prior for the binomial N.
673 *Ecology*, 94(10):2173–2179.

674 Link, W. A. and Barker, R. J. (2005). Modeling association among demographic
675 parameters in analysis of open population capture–recapture data. *Biometrics*,
676 61(1):46–54.

677 Liu, J. S. (1994). The collapsed {G}ibbs sampler in {B}ayesian computations with
678 applicationd to a gene regulation problem. *Journal of the American Statistical*
679 *Association*, 89(427):958–966.

680 Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS–A Bayesian
681 modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*,
682 10:325–337.

683 McClintock, B. T., Langrock, R., Gimenez, O., Cam, E., Borchers, D. L., Glennie, R., and
684 Patterson, T. A. (2020). Uncovering ecological state dynamics with hidden Markov
685 models. *Ecology letters*, 23(12):1878–1903.

686 Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978). Statistical
687 inference from capture data on closed animal populations. *Wildlife Monographs*,
688 62:3–135.

689 Piacenza, S. E., Balazs, G. H., Hargrove, S. K., Richards, P. M., and Heppell, S. S. (2016).
690 Trends and variability in demographic indicators of a recovering population of green sea
691 turtles *Chelonia mydas*. *Endangered Species Research*, 31:103–117.

692 Pledger, S., Baker, E., and Scribner, K. (2013). Breeding return times and abundance in
693 capture–recapture models. *Biometrics*, 69(4):991–1001.

694 Pledger, S., Pollock, KH., and Norris, JL. (2010). Open capture-recapture models with
695 heterogeneity II. Jolly-Seber model. *Biometrics*, 66:883–890.

696 Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using
697 Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed*
698 *Statistical Computing*. Vienna, Austria.

Royle, J. and Dorazio, R. M. (2012). Parameter-expanded data augmentation for Bayesian analysis of capture-recapture models. *Journal of Ornithology*, 152(S2):521–537.

Royle, J. A. and Dorazio, R. M. (2008). *Hierarchical Modeling and Inference in Ecology*. Academic Press- Elsevier Ltd.

Royle, J. A., Dorazio, R. M., and Link, W. A. (2007). Analysis of multinomial models with unknown index using data augmentation. *Journal of Computational and Graphical Statistics*, 16(1):67–85.

Saracco, J. F. and Yackulic, C. B. (2023). Fast and flexible Bayesian Jolly Seber models and application to populations with transients. *bioRxiv*, page 2023.05.25.542303.

Schofield, M. R. and Barker, R. J. (2008). A unified capture-recpature framework. *Journal of Agricultural Biological and Environmental Statistics*, 13:458–477.

Schofield, M. R., Barker, R. J., Link, W. A., and Pavanato, H. (2023). Estimating population size: The importance of model and estimator choice. *Biometrics*.

Schwarz, C. J. and Arnason, A. N. (1996). A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics. Journal of the International Biometric Society*, pages 860–873.

Scott, S. L. (2002). Bayesian methods for hidden Markov models: Recursive computing in the 21st century. *Journal of the American Statistical Association*, 97(457):337–351.

Seber, G. A. F. (1965). A note on the multiple-recapture census. *Biometrika*, 52:249–259.

Simpson, D. P., Rue, H., Riebler, A., Martins, T. G., and Sørbye, SH. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32:1–28.

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

Van Dyk, D. A. and Jiao, X. (2015). Metropolis-Hastings within partially collapsed Gibbs samplers. *Journal of Computational and Graphical Statistics*, 24(2):301–327.

White, G. C., Kendall, W. L., and Barker, R. J. (2006). Multistate survival models and their extensions in Program MARK. *Journal of Wildlife Management*, 70(6):1521–1529.

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

Zhang, W., Chipperfield, J. D., Illian, J. B., Dupont, P., Milleret, C., de Valpine, P., and Bischof, R. (2023). A flexible and efficient Bayesian implementation of point process models for spatial capture–recapture data. *Ecology*, 104(1):e3887.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R*. CRC Press, 2 edition.

Table 1: Parameter estimates for random effect *per capita* recruitment model for the European dipper data. Credible intervals are 95% highest probability density (HPD) intervals.

Parameter	Estimate	SD	95% CI
λ	324.67	24.21	[279.80 – 373.47]
N	324.61	16.25	[300 – 356]
f_1	1.27	0.76	[0.30 – 2.66]
f_2	0.70	0.19	[0.39 – 1.08]
f_3	0.53	0.10	[0.35 – 0.74]
f_4	0.48	0.09	[0.31 – 0.65]
f_5	0.49	0.08	[0.34 – 0.66]
f_6	0.43	0.08	[0.27 – 0.59]
ϕ_1	0.59	0.07	[0.46 – 0.75]
ϕ_2	0.51	0.06	[0.40 – 0.62]
ϕ_3	0.52	0.05	[0.42 – 0.61]
ϕ_4	0.59	0.05	[0.50 – 0.69]
ϕ_5	0.58	0.05	[0.50 – 0.67]
ϕ_6	0.58	0.06	[0.48 – 0.70]
p_1	0.64	0.27	[0.22 – 0.96]
p_2	0.82	0.10	[0.60 – 0.96]
p_3	0.89	0.05	[0.79 – 0.99]
p_4	0.89	0.05	[0.79 – 0.98]
p_5	0.89	0.04	[0.81 – 0.97]
p_6	0.91	0.04	[0.83 – 0.98]
p_7	0.88	0.07	[0.75 – 1.00]
μ_f	-0.55	0.28	[-1.04 – 0.03]
σ_f	0.48	0.34	[0.00 – 1.09]
μ_ϕ	0.25	0.19	[-0.10 – 0.63]
σ_ϕ	0.28	0.22	[0.00 – 0.68]
μ_p	1.85	0.55	[0.71 – 2.87]
σ_p	0.84	0.66	[0.02 – 2.03]

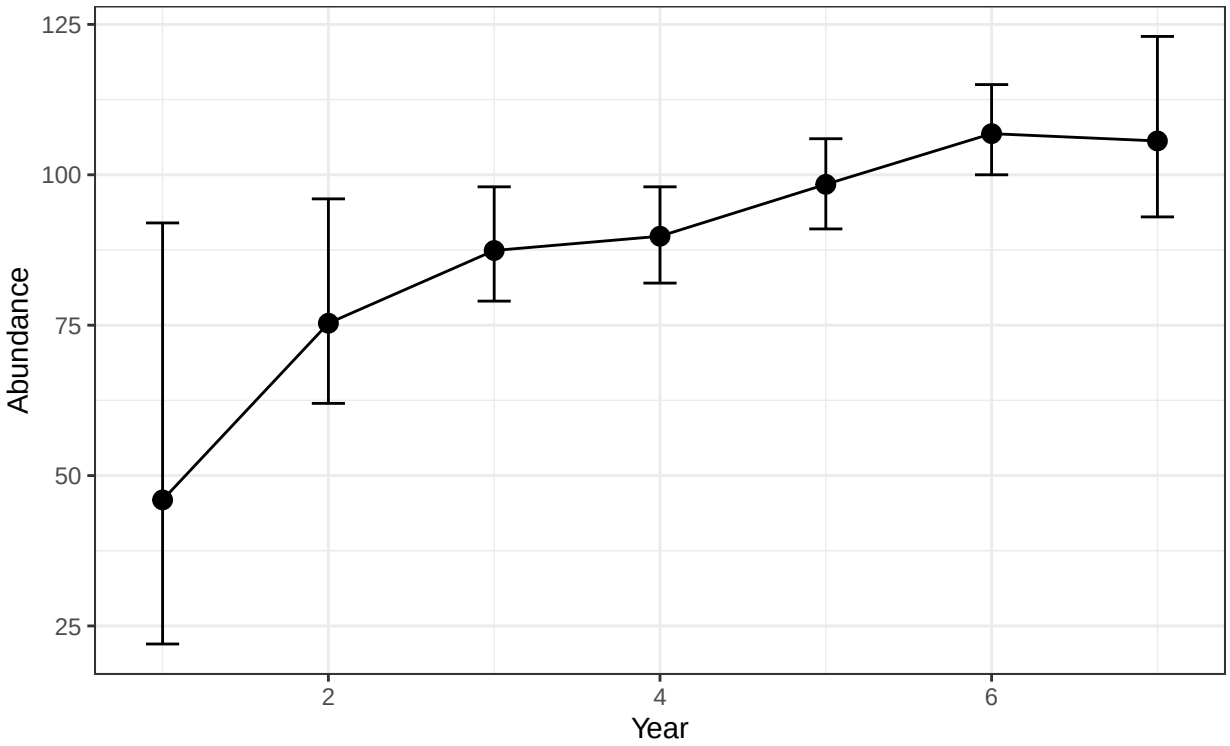

Figure 1: Time-specific abundance estimates for the European dipper data analysis. Error bars are 95% highest probability density intervals.

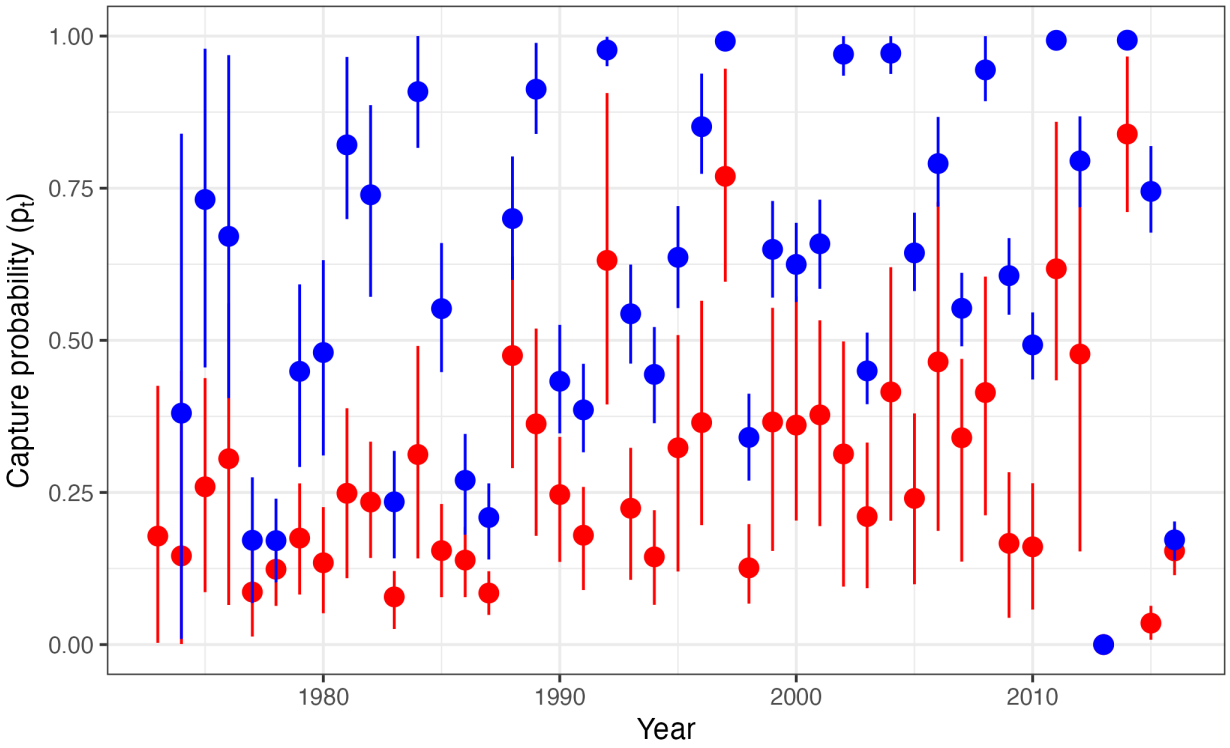

Figure 2: Capture probability estimates for Lalo honu data. Blue estimates are for individuals that have been previously captured (p_t)and red color denotes individuals that have not been previously captured ($\eta_t p_t$). Error bars are 90% highest probability density intervals.

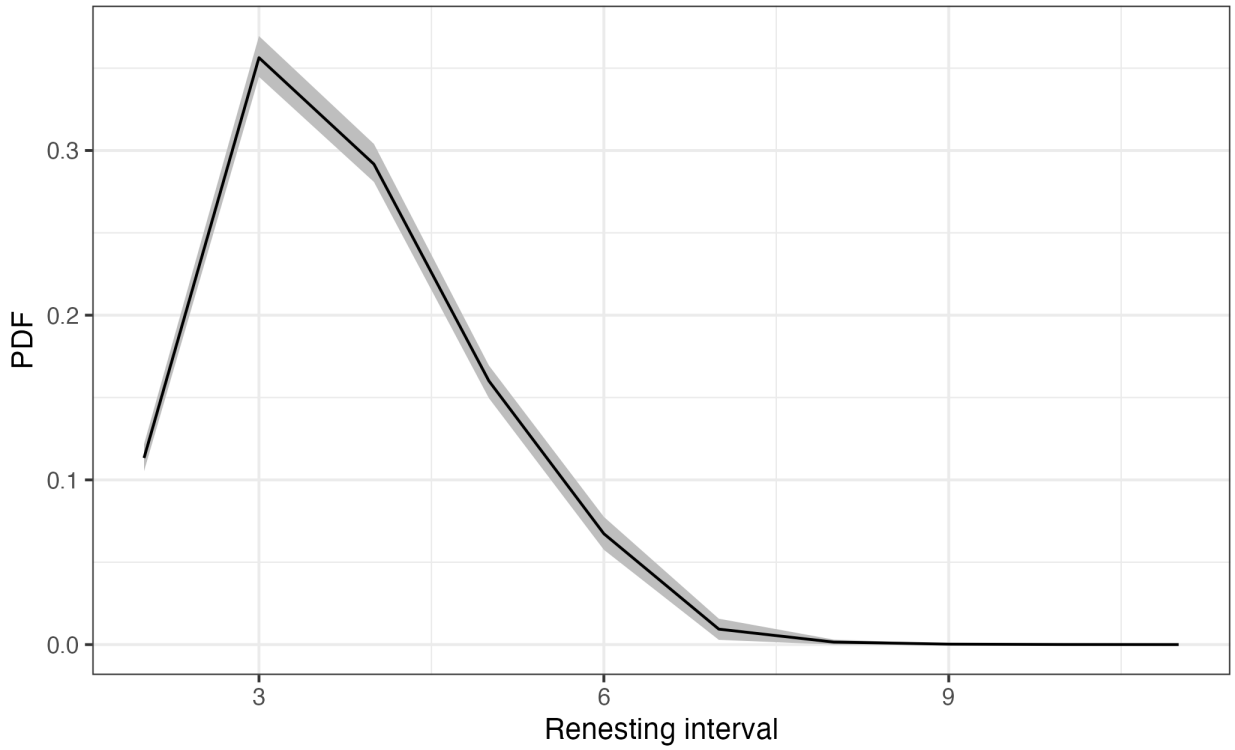

Figure 3: Estimated renesting interval probability distribution function for Lalo honu. Gray envelope is the 90% highest probability density interval.

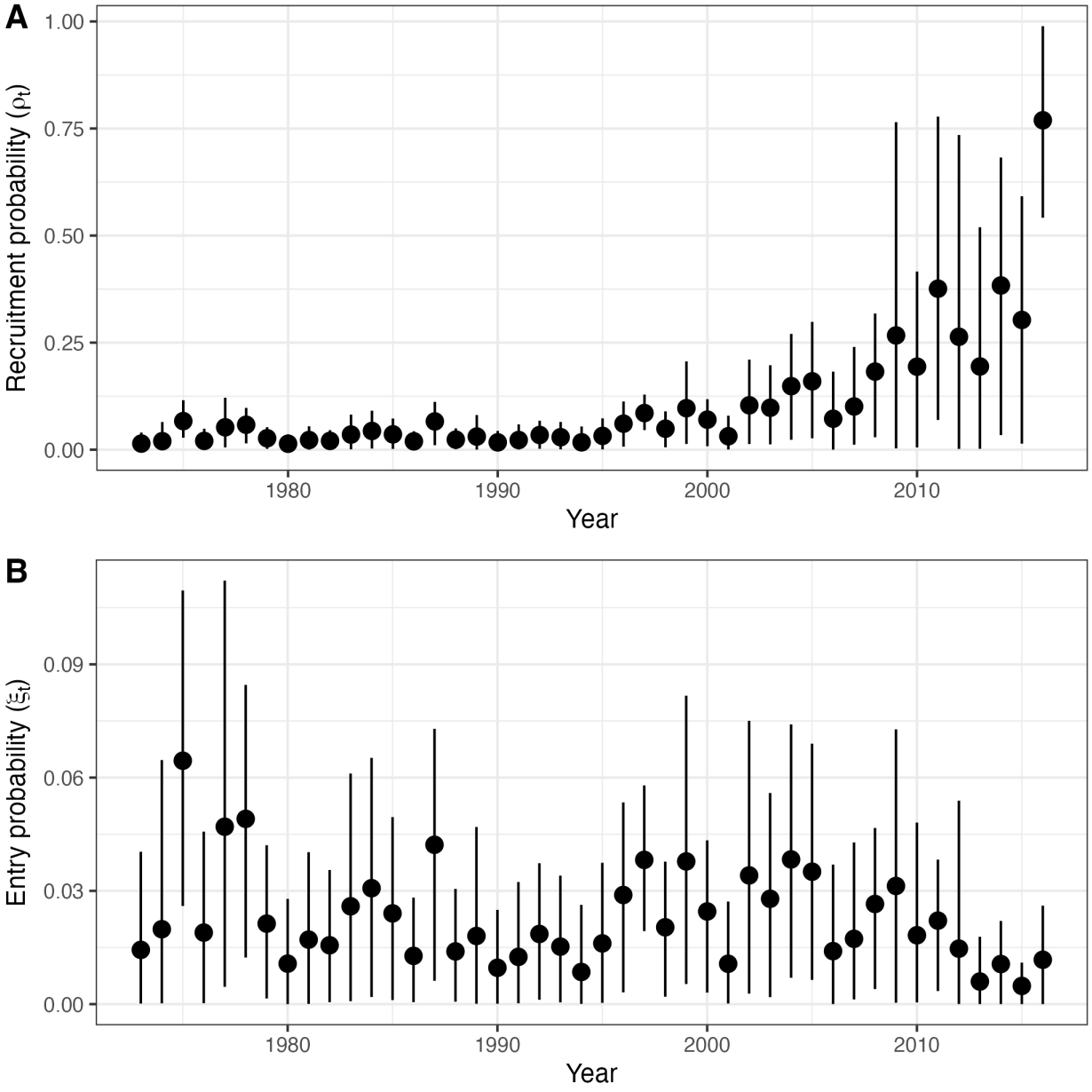

Figure 4: Plot A illustrates individual recruitment probability estimates (ρ_t) for Lalo honu. Plot B illustrates unconditional entry probabilities (ξ_t). In both plots error bars represent 90% highest probability density intervals.

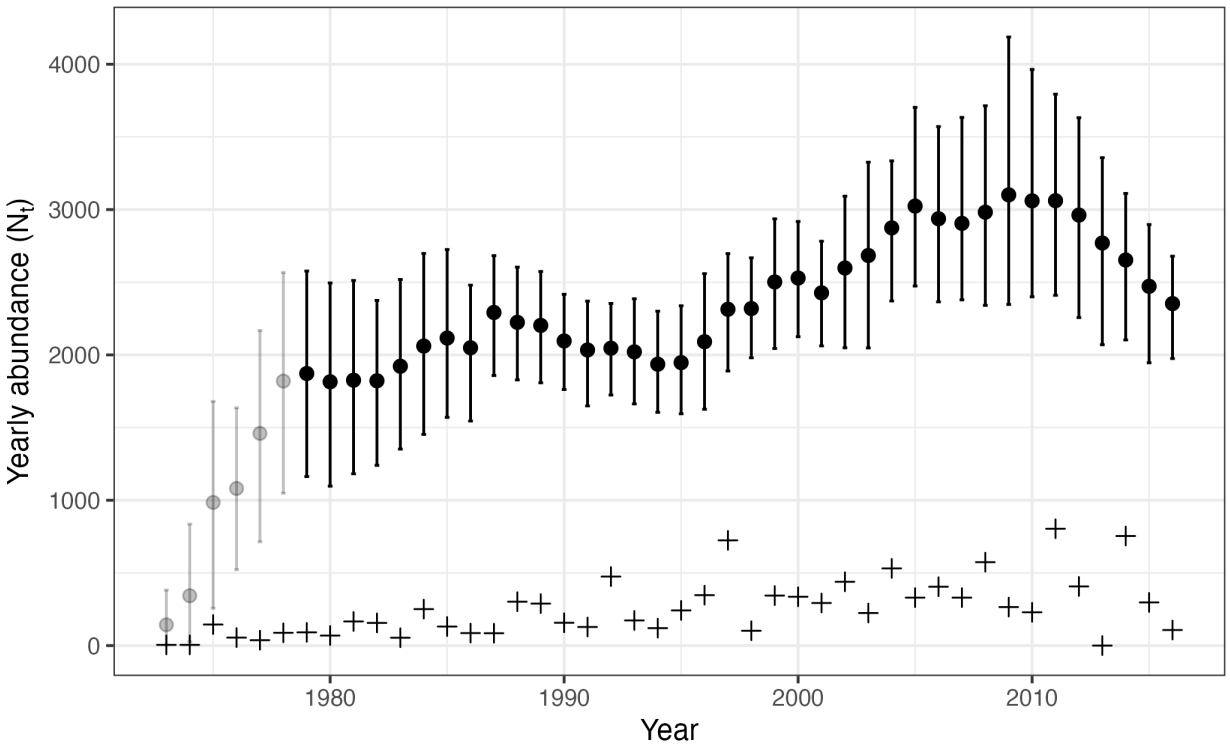

Figure 5: Estimated abundance of nesting female honu. The crosses are the number of observed individuals in a given year and the interval points are the estimated abundance in the entire population. The gray interval points are estimated abundance prior to the occasion when all individuals had a chance to be captured. Prior to that time there are some adult honu that were foraging and had not revisited Lalo for capture to be possible. Error bars represent 90% highest probability density intervals.

A Forward-Backward Posterior Abundance Sampling Algorithms

Here we present two versions of the Forward–Filtering Backward–Sampling (FFBS) algorithm for sampling the conditional posterior distribution of the true state of an individual(s). These algorithms are used within each round of MCMC updates to draw from the state distributions conditional on the observed data, $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, and the parameters, $\boldsymbol{\theta}$ at the current iteration. The first version is for updating the true state, $\mathbf{z}_i = (z_{i1}, \ldots, z_{iK})$ of an observed individual with capture history $\mathbf{x}_i \neq \mathbf{0}$. The second version is for updating the state and time-specific abundance of all unobserved individuals given the observed data, $\mathbf{x}$, the number of unobserved individuals, n_u, and the current parameters, $\boldsymbol{\theta}$. Both of these algorithms are coded as `nimble` functions (`sample_det_**` and `sample_undet_**`) in the `nimbleJSextras` package available at https://github.com/dsjohnson/nimbleJSextras or https://dsjohnson.r-universe.dev/nimbleJSextras.

Algorithm 1 FFBS for Observed Individuals

Inputs: $\mathbf{x}_i$, $\boldsymbol{\pi}$, $\{\boldsymbol{\Gamma}_t\}$, $\{\mathbf{P}_t(x_{it})\}$

Output: posterior draw of hidden state sequence $\mathbf{z}_i \sim [\mathbf{z}_i|\mathbf{x}_i, \boldsymbol{\theta}]$

Forward pass:

Compute $\mathbf{f}_1 = \boldsymbol{\pi}\mathbf{P}_1(x_{i,1})$

Normalize $\mathbf{f}_1 = \mathbf{f}_1/\text{sum}(\mathbf{f}_1)$

for $t = 1$ to $K - 1$ **do**

  $\mathbf{f}_{t+1} = \mathbf{f}_t\boldsymbol{\Gamma}_t\mathbf{P}_{t+1}(x_{i,t+1})$

  Normalize $\mathbf{f}_{t+1} = \mathbf{f}_{t+1}/\text{sum}(\mathbf{f}_{t+1})$

end for

Backward sampling:

Sample $z_{iK} \sim \text{categorical}(\mathbf{f}_K)$

for $t = K - 1$ down to 1 **do**

  Compute $\mathbf{b} = \mathbf{f}_t \odot \boldsymbol{\Gamma}_t[, z_{i,t+1}]$ ($\odot$ is elementwise multiplication)

  Normalize $\mathbf{b} = \mathbf{b}/\text{sum}(\mathbf{b})$

  Sample $z_{it} \sim \text{categorical}(\mathbf{b})$

end for

Return $\mathbf{z}_i = (z_{i,1}, \ldots, z_{i,K})$

Algorithm 2 FFBS for Unobserved Individuals

Inputs: $\boldsymbol{\pi}$, $\{\boldsymbol{\Gamma}_t\}$, $\{\mathbf{P}_t(0)\}$, n_u.

Output: posterior draw of abundance of unobserved individuals $\mathbf{n}_u \sim [\mathbf{n}_u|\mathbf{x}, n, \boldsymbol{\theta}]$, where $\mathbf{n}_u$ is a $J \times K$ matrix of state- and occasion-specific abundance of unobserved individuals.

Forward pass:

Compute $\mathbf{f}_1 = \boldsymbol{\pi}\mathbf{P}_1(0)$

Normalize $\mathbf{f}_1 = \mathbf{f}_1/\text{sum}(\mathbf{f}_1)$

for $t = 1$ to $K - 1$ **do**

  $\mathbf{f}_{t+1} = \mathbf{f}_t\boldsymbol{\Gamma}_t\mathbf{P}_{t+1}(0)$

  Normalize $\mathbf{f}_{t+1} = \mathbf{f}_{t+1}/\text{sum}(\mathbf{f}_{t+1})$

end for

Backward sampling:

Sample $\mathbf{n}_u[, K] \sim \text{multinomial}(n_u, \mathbf{f}_K)$

for $t = K - 1$ down to 1 **do**

  for $l = 1$ to J **do**

    Compute $\mathbf{b} = \mathbf{f}_t \odot \boldsymbol{\Gamma}_t[, l]$ ($\odot$ is elementwise multiplication)

    Normalize $\mathbf{b} = \mathbf{b}/\text{sum}(\mathbf{b})$

    Sample $\mathbf{Z}[, l] \sim \text{multinomial}(\mathbf{n}_u[l, t + 1], \mathbf{b})$

  end for

  for $l = 1$ to J **do**

    $\mathbf{n}_u[l, t] = \text{sum}(\mathbf{Z}[l,])$

  end for

end for

Return $\mathbf{n_u}$