Efficient Bayesian implementations of capture-recapture models with Stan

Matthijs Hollanders*†‡

Capture-recapture (CR) methods are a mainstay of ecological statistics for estimating demographic parameters and population sizes in animal populations. The advent of Bayesian methods made complex hierarchical formulations accessible to practitioners, largely relying on conditional likelihood formulations with latent discrete parameters. However, modern gradient-based MCMC methods that are the gold standard for sampling-based Bayesian estimation do not accommodate discrete parameters and require they are marginalised from the models. Here, I provide an overview of modern CR methods with efficient implementation in Stan, a probabilistic programming language. Models are categorised as Cormack-Jolly-Seber, conditioned on first capture, and Jolly-Seber, additionally estimating the entry process, with robust design, multistate, and multievent extensions covered for each type. All 16 model types are constructed in continuous time, using mortality and transition rates instead of probabilities, to accommodate unequal survey intervals. A novel component of this work is to accommodate unequal survey interval lengths in the entry process of Jolly-Seber models, which has been largely ignored despite being routinely accounted for in the survival process. In our case study, accounting for unequal intervals yielded better fit to data and considerable differences in population size estimates, highlighting the sensitivity of derived quantities to unrealistic model assumptions. Log likelihood functions and Stan programs are provided for all model types. Likelihood function are overloaded to accommodate both time-varying and individual-by-time varying effects, with the former in particular leveraging the factorisability of models to gain considerable efficiency gain. Jolly-Seber models, additionally, include functions to compute derived quantities like population sizes and entries and exits from the population using the forward-backward sampling algorithm.

Keywords: capture-recapture, Bayesian, Stan, Hamiltonian Monte Carlo, marginalisation, Cormack-Jolly-Seber, Jolly-Seber, robust design, multistate, multievent, hidden Markov model

*matthijs.hollanders@gmail.com
†Quantecol, Ballina, NSW, Australia
‡College of Engineering, Science and Environment, University of Newcastle, Newcastle, NSW, Australia

Introduction

Capture–recapture (CR) methods are a cornerstone of ecological statistics, used to estimate demographic parameters and produce unbiased estimates of population size in wild animal populations. Their power lies in tracking individual fates over time, yielding partially observed individual time series due to imperfect detection. Repeated surveys in a study area, coupled with the identification of individuals through natural markings (e.g., pelage patterns) or artificial tags, produce detection matrices from which quantities such as survival, recruitment, and population size can be estimated. Since their inception over a century ago (Petersen, 1896), CR methods have undergone extensive methodological development. Beginning with "closed-population" methods, where repeat counts facilitate estimating population size under the assumption populations are closed to entries (births and immigration) and exits (deaths and emigration), modern methods are "open-population" where these assumptions are relaxed.

Frequentist approaches to inference for these models have historically dominated CR analysis, exemplified by the popularity of Program MARK (White & Burnham, 1999), which has been cited 7,205 times (24 June 2025, Web of Science). The rise of Bayesian methods during the "MCMC revolution" (Martin et al., 2024) made it increasingly attractive to fit CR models via Markov chain Monte Carlo (MCMC), aided by user-friendly BUGS-language software such as WinBUGS (Lunn et al., 2000), JAGS (Plummer, 2003), and NIMBLE (de Valpine et al., 2017). BUGS-style modelling greatly improved accessibility to custom models in ecology, in part because latent discrete states——such as alive/dead status in CR, occupancy states in site-occupancy models, and latent abundances in N-mixture models—are straightforward to implement using conditional likelihood formulations. While discrete parameters are convenient for specifying models and deriving quantities such as population sizes, they are often computationally inefficient and can lead to poor exploration of the posterior tails (Stan Development Team, n.d.). Marginal likelihood formulations, which integrate over the discrete states, tend to be more efficient and lead to more efficient exploration of the posterior distribution while still permitting inference on those states.

Modern MCMC algorithms leverage gradients of the log posterior density for much more efficient exploration of the posterior than traditional samplers. The No-U-Turn Sampler (NUTS, Hoffman & Gelman, 2014) variant of Hamiltonian Monte Carlo (Neal, 2011) is a state-of-the-art algorithm implemented in probabilistic programming languages (PPLs) such as Stan (Carpenter et al., 2017), PyMC (Abril-Pla et al., 2023), and NIMBLE (de Valpine et al., 2017). Another key advantage of these algorithms is the ability to diagnose problems in the posterior geometry via signals like divergent transitions, alerting the practitioner that the estimation procedure cannot be trusted to produce correct inference (Betancourt, 2018). However, because HMC requires differentiable parameters (Neal, 2011), it cannot directly accommodate the discrete latent states of ecological models. This limitation has slowed adoption of HMC-based tools in ecology, despite their clear advantages over traditional samplers (Monnahan et al., 2017). Posterior inference on latent states in marginalised models is possible but requires specialised algorithms such as the forward-backward algorithm (Zucchini et al., 2017)——techniques that remain inaccessible to many practitioners (but see Kellner et al. (2021) and Socolar & Mills (2023) for recent contributions for ecological models).

In this paper, I provide an overview of CR methods and present novel, efficient Stan implementations of their marginal likelihoods. CR models are treated in two categories: Cormack–Jolly–Seber (CJS) models, which condition on an individual's first capture, and Jolly–Seber (JS) models, which explicitly model the process by which individuals enter the study and thus facilitate inference on population size and other population-level quantities. Within each category, single-state, multistate, and multievent variants, with and without robust design, are covered.

Overview of implementations

Descriptions and Stan implements are provided for a range of open-population CR models, beginning with Cormack–Jolly–Seber (CJS) models, which condition on first capture, and extending them to robust design obsermation models and multistate (including multievent) configurations. I then describe Jolly–Seber (JS) models, which additionally model the entry process of individuals in the population into the studied (or marked) sample and allow estimation of recruitment and population size. In all open-population CR models, the unit of observation is the time series of detections of a marked individual, making the individual the level where the likelihood factorises. Therefore, the likelihood is computed at the level of the individual, which is also where leave-one-out cross validation (LOO-CV) is evaluated (Bürkner et al., 2021; Socolar & Mills, 2023).

For each configuration, Stan programs and individual-level log-likelihood functions are presented that handle both (1) survey-varying parameters and (2) individual-by-survey–varying parameters. The first form enables precomputation of terms and is considerably faster. All likelihood functions were validated using simulation-based calibration (SBC, Modrák et al., 2023) with CmdStan 2.38 and CmdStanR 0.9 (Gabry et al., 2025) in R 4.5.1 (R Core Team, 2025). All Stan programs, associated functions, and SBC are provided at github.com/mhollanders/cr-in-stan.

Basic CR methods assume survey interval lengths are equal between sampling sessions (i.e., weekly or monthly intervals). However, because unequal survey intervals are the rule rather than the exception in field studies, unequal sampling intervals are accommodated in all models by parameterising ecological processes in continuous time, using rates rather than probabilities. For example, mortality is modeled via a hazard rate h, with survival probability ϕ over the survey interval τ given by $\phi^\tau = e^{-h\tau}$. Modelling hazard rates has several advantages over survival probabilities, including time invariance under a log-link, which entails that covariate effects are not affected by the chosen survey interval, unlike logit-linear functions for survival probabilities (Ergon et al., 2018). More importantly, unequal intervals in multistate models cannot be accommodated without continuous time transition rate matrices ($\boldsymbol{Q}$) (Glennie et al., 2023). The likelihood still requires discrete time transition probability matrices ($\boldsymbol{P}$) which cannot be simply exponentiated by a time interval because individuals may transition back and forth between different alive states multiple times within that interval, whereas mortality can only occur once. A novel parameterisation is also introduced for the entry process in JS models that incorporates unequal intervals, an aspect rarely addressed in previous work.

All Stan programs return the individual-level log likelihoods, enabling Pareto smoothed importance sampling leave-one-out cross-validation (PSIS-LOO-CV) at the level of the individual (Socolar & Mills, 2024) via the `loo` package (Vehtari et al., 2025) and prior log densities for sensitivity checks using powerscaling via the `priorsense` package (Kallioinen et al., 2023). For JS models, the forward-backward sampling algorithm is used to recover latent survey-level population sizes ($\boldsymbol{N}$), entries into ($\boldsymbol{B}$) and exits from ($\boldsymbol{D}$) the population for each survey, as well as the super-population N_{super} (the total number of individuals that were alive and available for capture during the entirety of the study). In multistate and multievent JS models, these quantities are computed for each state.

Efficiency in models is further improved by:

1. Vectorised computations for the probability of no further detections after a given survey;
2. Sharing this term and terms for the probability of first detection on a given survey across all individuals with the same first and last capture;
3. In JS models with data augmentation (Dupuis & Schwarz, 2007), computing the log likelihood of augmented individuals only once when no individual effects are present.

Models

Cormack-Jolly-Seber

The CJS model is the canonical open-population CR model (Cormack, 1964; Jolly, 1965; Seber, 1965), where unique individuals indexed $i \in \{1, \ldots, I\}$ are detected during surveys $j \in \{1, \ldots, J\}$ conducted in a study area. Following each individual's survey of first detection $f \in \{1, \ldots, J-1\}$[1], individuals survive survey intervals $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{J-1})$ with survival probabilities $\boldsymbol{\phi} = \exp(-\boldsymbol{h}\boldsymbol{\tau})$, where the mortality hazard rates $\boldsymbol{h}$ can be modeled flexibly as a function of individual, survey, and survey-varying individual effects. In CR models the ecological process usually estimates *apparent* survival, as we cannot generally disentangle mortality from permanent emigration from the site. Conditional on being alive, individuals are detected with probabilities $\boldsymbol{p} = (p_1, \ldots, p_{J-1})$, where p_j is the detection probability during survey $j+1$, which can similarly be modeled flexibly (one caveat is that time-varying mortality rates and detection probabilities lead to a lack of identifability in the final survey (Lebreton et al., 1992)). The assumed data-generating process for an individual is thus:

[1] Because CJS models condition on first capture, individuals first captured on the final survey do not contribute to the likelihood.

$$\begin{aligned} z_j &\sim \text{Bernoulli}\left(z_{j-1}\phi_{j-1}\right) \\ y_j &\sim \text{Bernoulli}\left(z_j p_{j-1}\right) \end{aligned}, \qquad j \in \{f+1, \ldots, J\}, \tag{1}$$

where $\boldsymbol{z} = (z_f, \ldots, z_J)$ are the latent states indicating whether an individual is alive ($z = 1$ is alive, $z = 0$ is dead), with $z_f = 1$, and $\boldsymbol{y} = (y_{f+1}, \ldots, y_J)$ is the observed detection history ($y = 1$ is detected, $y = 0$ is not). The marginal likelihood of an individual detection history $\boldsymbol{y}$ is given as follows, where for brevity $\boldsymbol{q} = \boldsymbol{1} - \boldsymbol{p}$ are the probabilities of not being detected:

$$\pi\left(\boldsymbol{y} \mid \boldsymbol{\phi}, \boldsymbol{p}, f, l\right) = \prod_{j=f+1}^{l} \left[\phi_{j-1} q_{j-1} \left(\frac{p_{j-1}}{q_{j-1}}\right)^{y_j}\right] \chi_l, \tag{2}$$

where $l \in \{f, \ldots, J\}$ is the last survey of detection, the expression $q\left(\frac{p}{q}\right)^{y_j}$ ensures a detection increments detection probability p and a non-detection increments q, and χ_l is the probability of never observing the individual again after its last capture. Note that the likelihood still uses ϕ despite it being parameterised in terms of h in the models. Typically, $\boldsymbol{\chi} = (\chi_1, \ldots, \chi_{J-1})$ is computed recursively which is inefficient when J is large. A more efficient computation is presented as one minus the probability of detecting the individual again during one of $m \in \{l+1, \ldots, J\}$ surveys (i.e. the sum of mutually exclusive potential detection occasions) (pers. comm. Dalton Hance). That is, for a given survey of last capture we express:

$$\chi_l = 1 - \sum_{m=l+1}^{J} \left[\phi_l \left(\prod_{j=l+2}^{m} q_{j-1}\phi_{j-1}\right) p_{m-1}\right], \qquad j \in \{1, \ldots, J-1\}. \tag{3}$$

This approach avoids redundant computations and is vectorisable, making it efficient in Stan. Moreover, $\boldsymbol{\chi}$ can also be pre-computed and shared across individuals when no individual effects are included.

Robust Design

To improve parameter identifability and precision in estimates of mortality rates, surveys can be conducted following the “robust design” (Pollock, 1982), which entails performing several “secondary surveys” indexed $k \in \{1, \ldots, K_j\}$ within primary surveys (e.g., 2 consecutive daily surveys within monthly primary sessions). The approach was originally introduced to account for temporary emigration (i.e., unavailability for capture), a variant of multistate model, but always improves parameter estimates by segregating the ecological and observation processes. Variable number of secondary surveys can be conducted, including $K_j = 1$ for some primaries. The closure assumption implies that we assume no animals are leaving the population, either through death or emigration, within a primary occasion. In robust design CJS models, we can now estimate detection probabilities in the primary of first capture, as we only need to condition on the secondary of first capture $g \in \{1, \ldots, K_j\}$, leaving $K_f - 1$ remaining detection probabilities in f. Additionally, the last survival and detection parameters are now also identifiable. The assumed data-generating process of the observation model is modified from Equation 2 as:

$$y_{jk} \sim \text{Bernoulli}\left(z_j p_{jk}\right), \quad j \in \{1, \ldots, J\},\ k \in \{1, \ldots, K_j\}, \tag{4}$$

and the marginal likelihood for an individual is expressed as follows, where $\boldsymbol{y}$ and $\boldsymbol{p}$ are now $J \times K$ matrices of detections/non-detections and detection probabilities, respectively, for each individual:

$$\begin{aligned} \pi\left(\boldsymbol{y} \mid \boldsymbol{\phi}, \boldsymbol{p}, f, g, l\right) = & \prod_{k \neq g}^{K_f} q_{fk} \left(\frac{p_{fk}}{q_{fk}}\right)^{y_{fk}} \\ & \cdot \prod_{j=f+1}^{l} \left[\phi_{j-1} \prod_{k=1}^{K_j} q_{jk} \left(\frac{p_{jk}}{q_{jk}}\right)^{y_{jk}}\right] \chi_l, \end{aligned} \tag{5}$$

where

$$\chi_l = 1 - \sum_{m=l+1}^{J} \left[\phi_l \left(\prod_{j=l+2}^{m-1} \phi_{j-1} \prod_{k=1}^{K_j} q_{jk}\right) \left(1 - \prod_{k=1}^{K_m} q_{mk}\right)\right], \tag{6}$$

with $1 - \prod_{k=1}^{K_m} q_{mk}$ giving the probability of making at least one detection in primary occasion m.

Multistate

Multistate models are a popular variant of CR models where alive individuals are assigned to $s \in \{1, \ldots, S\}$ distinct states, such as disease states (uninfected/infected), age (juvenile/adult), or breeding status (non-breeder/breeder). Instead of yielding detection vectors (or matrices in robust design) consisting of 0s and 1s to indicate an individual's (non)detections, multistate data consist of 0s for non-detections and $y_{ij[k]} = s$ to denote a detection in a particular state. Mortality rates and detection probabilities can be estimated separately for each state, as well as transition rates between states (e.g., infection dynamics). The ecological process consisting of mortality and state transitions is determined by a stochastic matrix ($\boldsymbol{P}$), equivalent called transition probability matrix, which are computed from an infinitesimal generator matrix ($\boldsymbol{Q}$) as $\boldsymbol{P}_j = \exp\left(\boldsymbol{Q}\tau_j\right)$, where exp is the matrix exponential, to accommodate unequal survey intervals. Stochastic matrices are parameterised with probabilties where rows (or columns) sum to 1, and generator matrices are parameterised with rates on the off-diagonals and the negative sum of the off-diagonals on the diagonal, ensuring that all rows sum to 0. With two states (i.e., $S = 2$) where transitions can occur freely between both states, $\boldsymbol{Q}$ and $\boldsymbol{P}$ are 3×3 matrices, where the final absorbing state is the dead state:

$$
\begin{aligned}
\boldsymbol{Q} &= \begin{bmatrix} -(q_1 + h_1) & q_1 & h_1 \\ q_2 & -(q_2 + h_2) & h_2 \\ 0 & 0 & 0 \end{bmatrix} \\
\boldsymbol{P} &= \exp(\boldsymbol{Q}),
\end{aligned}
\tag{7}
$$

where $\boldsymbol{h} = (h_1, \ldots, h_S)$ are the state-specific mortality rates and $\boldsymbol{q} = (q_1, \ldots, q_T)$ are the transition rates between states (maximum of $T = S(S - 1)$ distinct rates). Note that the states of departure (state during survey $j - 1$) are in the rows and the states of arrival (state during survey j) are in the columns. Detection probabilities $\boldsymbol{p} = (p_1, \ldots, p_S)$ can be estimated by state and survey and are represented in another transition probability matrix $\boldsymbol{O}$, given for the $S = 2$ example:

$$
\boldsymbol{O} = \begin{bmatrix} p_1 & 0 & 1 - p_1 \\ 0 & p_2 & 1 - p_2 \\ 0 & 0 & 1, \end{bmatrix}
\tag{8}
$$

where the latent ecological state is in the rows (where the final row represents probabilities associated with being dead) and the observed state is in the columns (with the final column representing a non-detection). Our assumed data-generating process for an individual is now as follows, where the matrix subscript subsets the associated row of the stochastic matrices:

$$
\begin{aligned}
z_j &\sim \text{Categorical}\left(\exp\left(\boldsymbol{Q}\tau_{j-1}\right)_{[z_{j-1}]}\right) \\
y_j &\sim \text{Categorical}\left(\boldsymbol{O}_{[z_j]}\right),
\end{aligned}
\qquad j \in \{f + 1, \ldots, J\}
\tag{9}
$$

Note that the typical approach in BUGS/JAGS is to convert $y_{ij[k]} = 0$ to $S + 1$ and to model by the dead state explicitly, but this is not required when specifying the log density in Stan. The marginal likelihood is unwieldy to express (see Stan programs), but our provided Stan implementations use the forward algorithm (Zucchini et al., 2017) for alive states only (that is, omitting row and column $S + 1$ from $\boldsymbol{P}$) until last capture, limiting unnecessary computation depending on whether or not individuals were detected in occasions j and $j - 1$. Additionally, we compute $\boldsymbol{\chi}$, the probability of not making another detection after last being observed in a particular state, efficiently by modifying Equation 3 to compute the probability of making at least one more detection while accounting for possible state transitions. Robust design implementations are also provided, which are generally required to identify state-specific detection probabilities alongside transition rates (Hollanders & Royle, 2022). Our SBC results qualitatively reflect this, with large uncertainties associated with the single survey variants.

Multievent

Where multistate models assume that individuals are always correctly observed in their respective states (e.g., individuals infected by a pathogen are never detected without infection), multievent models account for imperfect state assignment or state uncertainty by including false negatives and/or positives (Pradel, 2005). Failing to account for incorrect state assignment can seriously bias parameter estimates, particularly transition rates between respective states, and multievent models should be the default in such circumstances (Hollanders & Royle, 2022). Multievent models are constructed like multistate models except that observed individuals have an additional $S \times S$ stochastic matrix $\boldsymbol{E}$ associated with their state assignments, shown here for $S = 2$ with false negatives $(1 - \delta)$ and false positives (λ) associated with being detected in state $S = 2$ (for example, in disease-structured models where state 2 is the infected state):

$$\boldsymbol{E} = \begin{bmatrix} 1 - \lambda & \lambda \\ 1 - \delta & \delta \end{bmatrix}. \tag{10}$$

The uncertain state assignment means that unlike in typical multistate models, the initial states at first capture are unknown. The initial states are governed by a simplex $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_S)$, where $\sum_j \eta_j = 1$, which may further vary by survey, $z_f \sim \text{Categorical}\left(\boldsymbol{\eta}_{[f]}\right)$. These parameters may not always be of direct ecological interest, and Stan can handle survey-varying simplices efficiently with the new `stochastic_matrix` parameter types. Alternatively, the stationary distributions can be computed from $\boldsymbol{q}$ to reduce model dimensionality, an approach that is also implemented for similar models in the R package `hmmTMB` (Michelot, 2025). When $S = 2$, the stationary distribution can be computed from the transition rates by normalising $\boldsymbol{q}$, potentially by survey j, and reversing the order of the elements,

$$\boldsymbol{\eta}_j := \frac{\text{reverse}\left(\boldsymbol{q}_j\right)}{\sum \boldsymbol{q}_j}. \tag{11}$$

Multievent models generally require robust design sampling for adequate parameter identifiability unless more informative priors are used (Hollanders & Royle, 2022), and models with both false negatives and false positives generally have a multi-modal likelihood (Royle & Link, 2006), potentially requiring some parameters constraints to imposed (e.g., $\lambda < \delta$). They are also computationally costly in Stan, losing a considerable degree of factorisability as potentially none of the observations can ever be assigned to a state with certainty. Additionally, the multievent constructions do not compute $\boldsymbol{\chi}$ but rather marginalise over all possible ecological states, including the absorbing dead state, use the forward algorithm. Nevertheless, when state assignment errors cannot be ruled out, as is commonplace is some study designs, multievent models are the appropriate way to produce unbiased estimates of the ecological process under study.

Jolly-Seber

JS models feature the same likelihood computations as CJS after first capture but additionally model the process by which individuals enter the study, which in some contexts may be interpreted as recruitment (Schwarz & Arnason, 1996). Observed individuals I are considered as a subset of the super-population N_{super}, defined as the total number of individuals that were ever alive and available for capture in the study. Each individual in the super-population must enter during one of J surveys with entry probabilities forming a simplex, with the entry survey modeled as:

$$b \sim \text{Categorical}\left(\boldsymbol{\beta}\right), \qquad \sum_{j=1}^{J} \beta_j = 1, \tag{12}$$

This renders the number of entries to the population during each survey $\boldsymbol{B} = (B_1, \ldots, B_J)$ multinomially distributed as $\boldsymbol{B} \sim \text{Multinomial}\left(N_{\text{super}}, \boldsymbol{\beta}\right)$.

The entry probabilies here have an ecological interpretation and can be modeled flexibly, although the interpretation of β_1 differs fundamentally, as it represents the proportion of the superpopoulation that is the starting population size at $j = 1$. This contrasts with the classic approach employed in BUGS, where $\boldsymbol{\beta}$ is generally derived from the ecologically uninterpretable "removal entry probabilities" (Dorazio, 2020; Royle & Dorazio, 2012).

In JS models there are J detection probabilities to estimate, as the unknown entry occasion $b_i \in \{1, \ldots, J\}$ means we do not condition on first capture. The data-generating process for an individual in a JS model is thus as follows:

$$\begin{aligned} b &\sim \text{Categorical}\left(\boldsymbol{\beta}\right) \\ z_j &\sim \text{Bernoulli}\left(z_{j-1}\phi_{j-1}\right), && j \in \{b+1, \ldots, J\} \\ y_j &\sim \text{Bernoulli}\left(z_j p_j\right), && j \in \{b, \ldots, J\} \end{aligned} \tag{13}$$

In the likelihood computation we have to marginalise out the discrete entry occasions $\boldsymbol{b}$, which involves incrementing the following term to Equation 2:

$$\sum_{b=1}^{f} \left[\beta_b q_b \prod_{j=b+1}^{f} \phi_{j-1} q_j \right] \frac{p_f}{q_f} \tag{14}$$

In the absence of individual effects, individuals with the same survey of first capture share the same likelihood prior to first capture, reducing computational burden. JS models also benefit from the robust design as it leads to identifiability of the first entry and detection probabilities (Schwarz & Arnason, 1996). Stan programs and associated functions are provided for robust designs, and the SBC results clearly demonstrates the increase in precision of the parameter estimates compared to survey designs without robust design (see GitHub).

Since N_{super} is unknown, data augmentation (Dupuis & Schwarz, 2007) of I_{aug} all-0 (never detected) detection histories is used with a nuisance inclusion parameter ψ to determine what proportion of the total capture histories belonged to real individuals in N_{super}, where $N_{\text{super}} \sim \text{Binomial}\left(I + I_{\text{aug}}, \psi\right)$. The Stan functions included on GitHub (specifically the `js_*_rng()` functions that compute derived quantities) print warnings when $N_{\text{super}} = I + I_{\text{aug}}$ for each posterior draw, indicating the detection history was insufficiently augmented with all-0 detection histories.

A key advantange of using Stan to specify the log likelihood is that when there are no individual effects, the log likelihood is the same for all augmented individuals and can thus be computed once, greatly reducing computational burden. For JS models, additional likelihood function signatures are provided (e.g., `js2()`) that accommodate individual effects for the observed sample but only perform one computation for augmented individuals. This "collapsed" likelihood version assumes all augmented individuals share the same parameters (i.e., the average across the sampled population), and forego imputing values for each augmented individual, therefore reducing computational burden (Royle, 2009). Practitioners could use these functions to determine roughly how many all-0 detection histories should be augmented before running a "full" model where parameters are imputed for all augmented individuals, to avoid computing a large proportion of log likelihoods for surplus augmented individuals.

Parametric forms for the entry probabilities

The canonical distribution for modeling simplexes is the Dirichlet such that $\boldsymbol{\beta} \sim \text{Dirichlet}\left(\boldsymbol{\alpha}\right)$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)$ is the concentration (or shape) vector which in the context of CR may be loosely interpreted as the "entry rates" or expected number of individuals entering per survey. The Dirichlet distribution is produced by normalising the vector $\boldsymbol{u} \sim \text{Gamma}\left(\boldsymbol{a}, \mathbf{1}\right)$ (with $\mathbb{E}(\boldsymbol{u}) = \boldsymbol{\alpha}$) where $\beta_j := \frac{u_j}{\sum_{j=1}^{J} u_j}$, thus it can be considered a normalised Gamma process. Typically, entry probabilities are modeled as a symmetric Dirichlet distribution, where $\alpha_j = \mu$, and estimating μ as a parameter is possible to determine the "sparsity" of the entry probabilities. When $\mu = 1$, the simplex is uniform; when $\mu < 1$, the simplex is more sparse (some surveys have very low entry probabilities while others have high entry probabilities); and when $\mu > 1$, the simplex gets pulled closer toward equal entries per survey. Although the Dirichlet is a natural approach for modelling the number of entries in a multinomial process, the covariance structure is completely determined by the vector $\boldsymbol{\alpha}$ and is always negative between the different entries of the vector (Aguilar & Bürkner, 2025).

As an alternative to the Dirichlet, the logistic-normal distribution (Aitchison & Shen, 1980) can be used to model the simplex of entry probabilities with a flexible covariance structure. Here, $\beta_j := \frac{e^{v_j}}{\sum_j^J e^{v_j}}$ where $\boldsymbol{v} \sim \text{Normal}\left(\mathbf{0}, \Sigma\right)$ and can be considered a normalised lognormal process, where $\frac{e^{v_j}}{\sum_j^J e^{v_j}}$ is the softmax transformation. Since the softmax function is translation invariant (i.e., unchanged by adding the same constant to every component of the vector), only $J - 1$ elements are freely estimable. Typical approaches are to fix one element of $\boldsymbol{v}$ to 0 or to impose a sum-to-zero constraint; in Stan, the `sum_to_zero_vector` parameter type makes implementing the latter approach trivial. The sparsity of the entry probabilities here is determined by the covariance matrix: if Σ is a diagonal matrix $\mu \boldsymbol{I}_J$, then larger values of μ support a sparse simplex while lower values of μ produces a simplex more concentrated around the expected entry probabilities. Note that in the Dirichlet case, a sparse simplex is produced by *smaller* values of μ.

The provided JS Stan programs on GitHub accommodate both entry processes where an indicator is supplied as data to use the Dirichlet (`dirichlet = 1`) or logistic-normal (`dirichlet = 0`) as a prior for the entry process. For the simulations conducted for SBC, the logistic-normal was used as it is considerably faster. Although the Dirichlet is a natural choice, and indeed the conjugate prior for the number of entries, the logistic-normal facilitates directly modeling more complex covariances, such as temporal correlations in the entry probabilities via a Gaussian process (Case Study 2).

Accommodating unequal intervals in the entry process

A novel contribution of this manuscript is to accommodate unequal survey intervals into the entry process. Analogous to how mortality rates are multiplied by the survey intervals, the intervals can also be incorporated into expectations of the entry process, where for the Dirichlet we can model the concentration vector as multiplied by the survey interval:

$$
\begin{aligned}
\boldsymbol{\beta} &\sim \text{Dirichlet}\left(\boldsymbol{\alpha}\right) \\
\alpha_1 &= \mu\gamma \\
\alpha_j &= \mu\tau_{j-1}, \qquad\qquad j \in \{2, \ldots, J\}
\end{aligned}
\tag{15}
$$

Here, μ controls the overall concentration (or “entry rate” per unit of the survey interval length) of the Dirichlet prior, while the relative weights of α_j are determined by the survey intervals τ_j. The parameter γ scales μ for the first survey, as this entry occasion has no associated time interval. The parameterisation is most interpretable when $\boldsymbol{\tau}$ has been scaled to have geometric mean of 1 (Box 1). In that case, when $\gamma > 1$, the starting population is large relative to the super-population, and vice versa when $\gamma < 1$.

The logistic-normal case can be written as follows, where I use the same symbols as in the Dirichlet model for the corresponding distribution components, where in this case the covariance matrix Σ is a diagonal matrix with μ for each diagonal entry:

$$
\begin{aligned}
\boldsymbol{\beta} &= \frac{e^{\boldsymbol{v}}}{\sum_{j=1}^{J} e^{v_j}} \\
\boldsymbol{v} &\sim \text{Normal}\left(\log \boldsymbol{\alpha}, \mu \boldsymbol{I}_J\right), & \sum_{j=1}^{J} v_j = 0 \\
\alpha_1 &= \gamma \\
\alpha_j &= \tau_{j-1}, & j \in \{2, \ldots, J\}
\end{aligned} \tag{16}
$$

Again, when $\boldsymbol{\tau}$ is scaled to have a geometric of 1, $\gamma > 1$ indicates a large starting population.

Multistate and multievent JS models

Multistate and especially multievent JS models are computationally demanding because the likelihood needs to not only be marginalised over possible entry occasions, but also possible entry states. Unlike multistate CJS models, where we condition on first capture and assume states are perfectly observed, in multistate JS models we do not know the entry occasion, and thus also not the state at the time of entry. For example, in a multistate model with $S = 3$ the likelihood for individual first captured on $j = 4$ requires marginalising over 4 possible entry occasions with 3 different entry states each, requiring some part of the likelihood computations to be performed 12 times. The entry process computations can be pre-computed in the absence of individual effects, greatly increasing speed. Multievent models, which incorporate no certain states by which to factor the different likelihood terms, require the log likelihoods for the individual's entire detection vector to be computed for all possible entry occasions and entry states. Such models may become prohibitively expensive to compute for large datasets, though given that the likelihood terms are computed for each individual, Stan's within-chain parallelisation options will be beneficial to share likelihood computations across cores. Again, preliminary model runs without individual effects will offer great efficiency gains to determine the required number of augmented individuals before running full models with all desired effects.

Case studies

Two example datasets were chosen to showcase the CR implementations in Stan.

Case Study 1: Endangered Houston toads

I fit a JS model to a previously analysed dataset of endangered Houston toads (Duarte et al., 2011) with three different priors for the entry probabilities $\boldsymbol{\beta}$:

1. Uniform Dirichlet prior, $\boldsymbol{\beta} \sim \text{Dirichlet}\,(\mathbf{1})$;
2. Dirichlet prior with concentration parameter estimated, $\boldsymbol{\beta} \sim \text{Dirichlet}\,(\mu\mathbf{1})$, where μ was given a $\text{Gamma}\,(1, 1)$ prior;
3. The same as (2) but with unequal intervals incorporated, using a $\text{Gamma}\,(1, 1)$ prior for γ (Equation 15).

I compared models using expected log predictive density (ELPD) as implemented in the loo R package (Vehtari et al., 2025). The uniform Dirichlet prior performed considerably worse, with ELPD being >3 standard errors lower. Model (3) incorporating interval lengths had the highest ELPD, but not significantly compared to model (2), which only estimated a concentration parameter. The models supported that entry probabilities formed a "sparse simplex", with few surveys accounting for the majority of entered individuals. This, in turn, had a considerable impact on the estimated population sizes, with model (1) producing lower maximum population sizes compared to the other models (Figure 1).

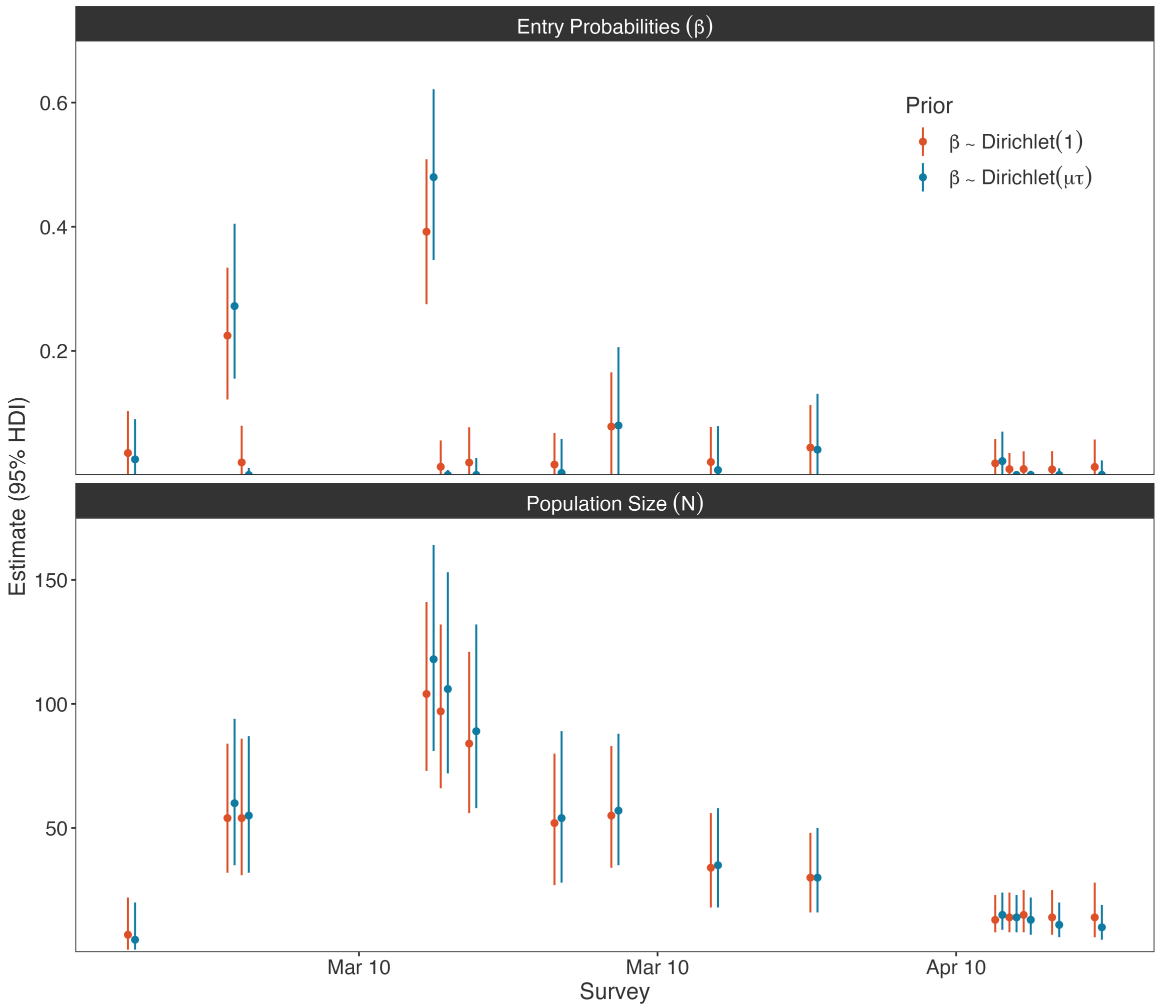

Figure 1: Entry probabilities ($\boldsymbol{\beta}$) and population sizes ($\boldsymbol{N}$) of Houston toads from Jolly-Seber models (summarised with posterior medians and 95% HDIs) with two different priors for entry probabilities. The prior incorporating unequal survey intervals along with an estimated Dirichlet concentration parameter (μ) performed best according to expected log predictive density (ELPD). Data came from Duarte et al. (2011).

Case Study 2: Contemporary response to the amphibian chytrid fungus

I fitted a robust design multistate JS model to a dataset of Fleay's barred frogs (*Mixophyes fleayi*), an endangered Australian frog for which the contemporary response to the pathogenic amphibian chytrid fungus (*Batrachochytrium dendrobatidis*, *Bd*) was originally assessed with a CJS model (Hollanders et al., 2023). The objective was to determine differential mortality rates between $S = 2$ states, frogs uninfected and infected with the fungus, and data were collected from 587 frogs at 3 sites (Brindle, Tuntable, and Bat Cave Creeks) across 21 primary occasions (with 2–3 secondaries per primary) spanning 4 years, yielding 2363 total detections. Only males were included in the current analysis due to the paucity of female recaptures. Site-level intercepts for mortality and transition rates and detection probabilities were modeled hierarchically using zero-sum effects for identifiability (Ogle & Barber, 2020), and detection probabilities were assumed to not have been affected by infection status.

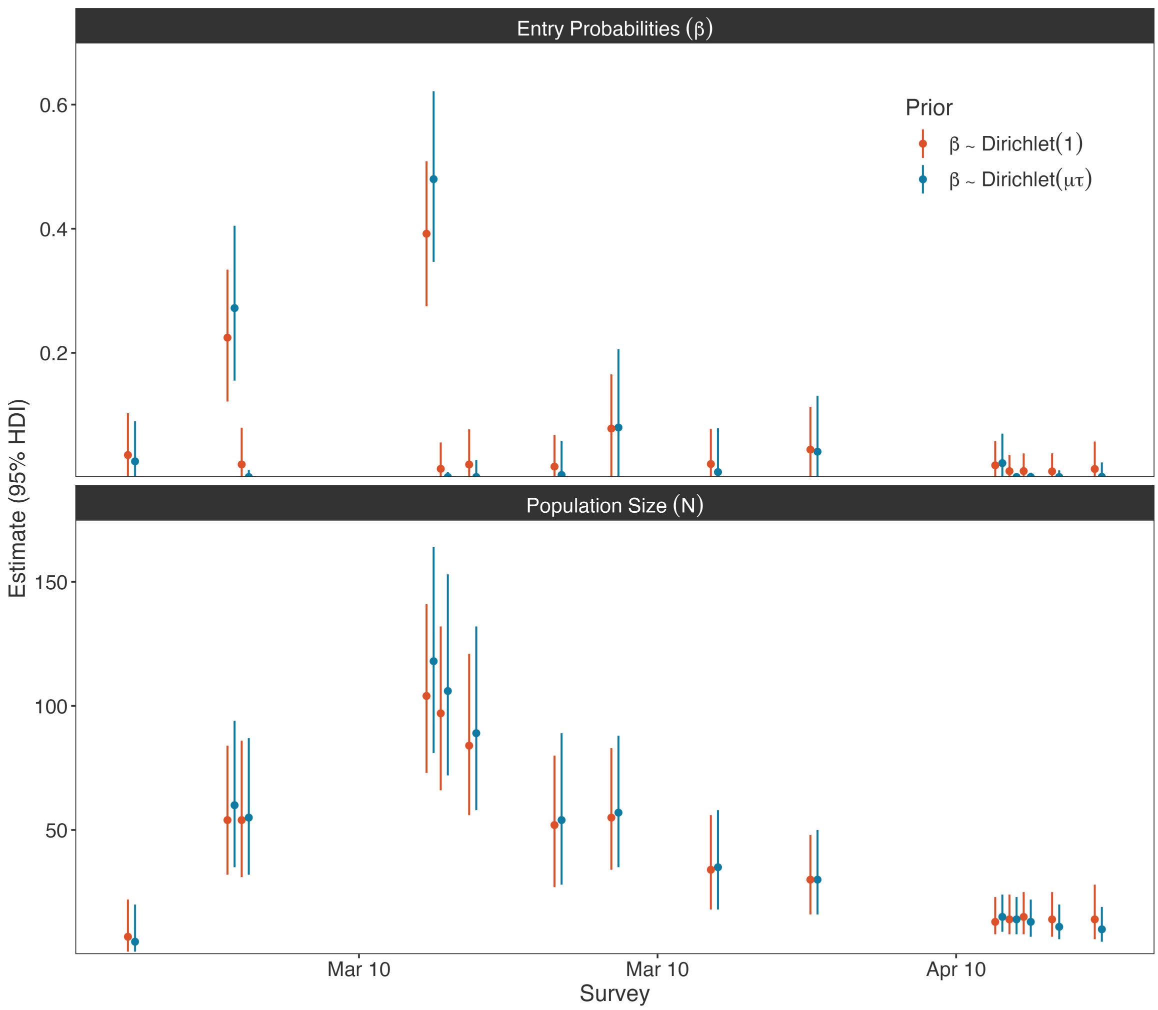

Figure 2: Fleay's barred frog (*Mixophyes fleayi*) population sizes and infection prevalence, computed as the proportion of individuals estimated to be infected with *Bd*, derived with forward-backward sampling of latent states (summarised with posterior medians and 95% HDIs). Data from Hollanders et al. (2023).

A multivariate Gaussian processes (GP) was implemented to capture temporal heterogeneity in 5 parameters: mortality (h, shared across uninfected and infected frogs), infection dynamics ($\boldsymbol{q}$, rates of gaining and clearing Bd), detection probabilities (p), and entry probabilities (β, thus resulting in a logistic-normal prior) (Box 1). To reduce model dimensionality, we parameterised the initial state probabilities ($\boldsymbol{\eta}$) as the stationary distribution from infection dynamics (Equation 11). The multivariate GP was modeled with both row and column covariances. The row covariance was determined by a $J \times J$ exponentiated quadratic kernel ($\boldsymbol{K}$) using the survey dates as input, with the the marginal scale fixed to 1 for identifability, and using a single length-scale to capture temporal correlation (Box 1). The column covariance was determined by a 5×5 variance-covariance matrix (Σ), facilitating the estimation of the temporal variance in each parameter as well as a correlation matrix for the parameters. The GP realisations were then modeled as matrix normal $f \sim \text{MatrixNormal}\left(\mathbf{0}, \boldsymbol{K}, \Sigma\right)$, which was implemented using the non-centered parameterisation in Stan as $f = \boldsymbol{L}\boldsymbol{z}\boldsymbol{U}$, where $\boldsymbol{L}$ is the lower triangular Cholesky factor of $\boldsymbol{K}$, $\boldsymbol{U}$ is the upper triangular Cholesky factor of Σ, and $\boldsymbol{z}$ is a $J \times 5$ matrix of standard normal variates. The Stan program used is available in the GitHub repository.

We ran 8 chains for 500 iterations after 500 warm-up draws which was sufficient for convergence with <1% divergent transitions, which took 12.3 mins on an M2 Macbook Pro. The analysis revealed that mortality rates between uninfected and infected individuals are similar, resulting in mostly stable population sizes over the study period despite peaks of infection during cooler months, although the Bat Cave Creek population appears to be declining (Figure 2). Another result was that temporal mortality rates and entry rates were strongly positively correlated (0.77, 95% HDI 0.45, 0.96), perhaps indicating increased movement in warmer months, resulting in entries and exits from the population. This would imply that the increase in apparent mortality is driven by permanent emigration, not mortality (Figure 3). The result is further corrobated by a positive correlation between entry rates and detection probabilities (0.36, 95% HDI 0.02, 0.66). During periods of higher activity, frogs have higher detectability and also move around move, migrating into and out of the streamside transects comprising the study area.

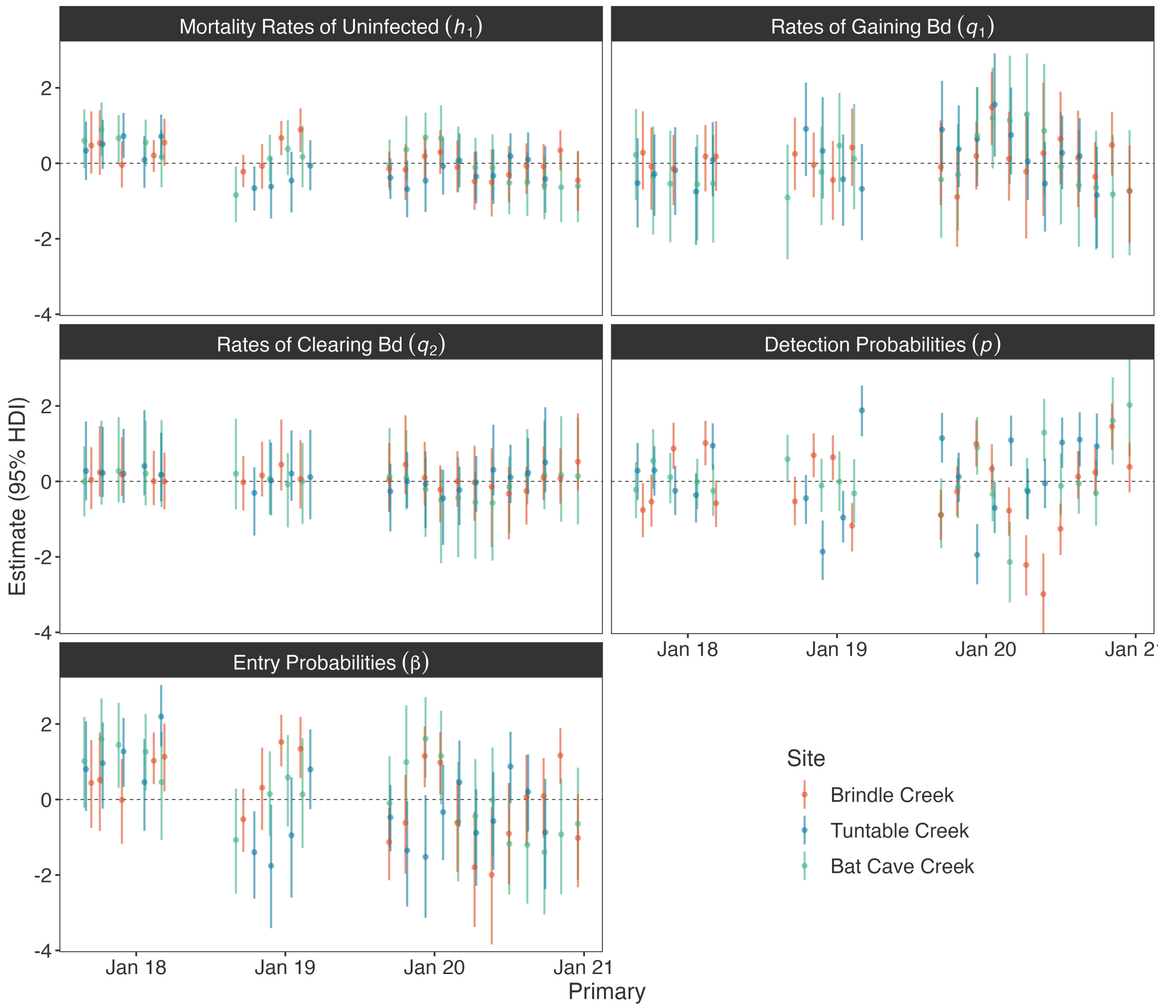

Figure 3: Multivariate Gaussian process (GP) realisations for each model parameter, centred on 0 (summarised with posterior medians and 95% HDIs). A strong positive correlation between mortality and entry rates (0.77, 95% HDI 0.45, 0.96) may suggest increased apparent mortality is indicative of dispersal, leading to permanent emigration.

Discussion

A range of CR models were reviewed with efficient implementations of all types provided in Stan. As gradient-based MCMC methods demand discrete parameters are marginalised from the models, these state-of-the-art sampling algorithms remained largely inaccessible for complex ecological models to most practitioners. Moreover, in order to recover posterior distributions of latent discrete parameters to estimate quantities like population size, the forward-backward sampling algorithm is required in addition to the likelihood computation. This contribution provides functions for computing the log likelihoods and recovering latent quantities for a range of models. Since these models are implementations of hidden Markov models which are statistically equivalent to other ecological models such as dynamic occupancy models (McClintock et al., 2020), these functions can readily be modified for a range of other applications.

Since many CR models can be factored as product multinomials by conditioning on first and last detections, some likelihood terms can be shared across individuals in the absence of individual-level effects. This greatly reduces computational burden, especially for JS models with data-augmentation where the likelihood of an augmented individual need only be computed once. The factorisability of the likelihood is reduced with the more complex model variants such as multistate and especially multievent, as capture events can no longer be associated with fixed underlying ecological states. In JS models with categorical predictors, there are addditional complications, where the likelihood terms of augmented individuals have to be computed for the combinatorics of possible categorical states. With sex, for instance, augmented individual likelihoods have to be computed for males and females and then weighted according to the sex ratio, which may itself be a parameter to be estimated.

Unequal survey intervals have typically only been accounted for in the survival and multistate transition processes. In the simplest single state models, exponentiating the survival probability by the survey interval is sufficient, but this approach does not work for multistate models (Glennie et al., 2023). Therefore, a more general approach is to multiply the hazard rates (or generator matrix in multistate) by the survey interval followed by taking the (matrix) exponential to produce survival and transition probabilities, respectively. This paper further extends this idea to the entry process, where longer survey intervals are expected to produce relatively more entries to the population. Implementing the approach described herein produces considerably different latent population sizes and entry probabilities, with higher ELPD via LOO-CV model comparison, highlighting the importance of accommodating varying survey intervals (Figure 1).

Despite statistical advances in estimating wild animal population sizes, methods relying on unmarked populations are highly sensitive to model assumptions (Barker et al., 2018; Gilbert et al., 2021), cementing CR as the de facto method for estimating demographic processes in animals. Advances in Bayesian computational methods now rely on gradient-based MCMC methods, which require the marginal likelihoods to be fitted. Fittingly, these same marginal likelihood constructions were used in the frequentist paradigm that has dominated the CR literature. This manuscript bridges the gap between cutting edge Bayesian model fitting and fundamental ecological statistical methods.

Data and Code Availability

All Stan programs and associated functions, as well as data and programs used in the case studies, are available at github.com/mhollanders/cr-in-stan.

Acknowledgements

I am grateful to Dalton Hance for efficiently computing the probability of no more observations after last capture (χ), Bob Carpenter to providing guidance on accounting for unequal survey intervals in the entry process, Marc Kéry to feedback on the manuscript, and Ben Augustine for helpful discussions.

References

Abril-Pla, O., Andreani, V., Carroll, C., Dong, L., Fonnesbeck, C. J., Kochurov, M., Kumar, R., Lao, J., Luhmann, C. C., Martin, O. A., et al. (2023). PyMC: A modern, and comprehensive probabilistic programming framework in Python. *PeerJ Computer Science*, *9*, e1516.

Aguilar, J. E., & Bürkner, P.-C. (2025). *Generalized Decomposition Priors on R2* (arXiv:2401.10180). arXiv. https://doi.org/10.48550/arXiv.2401.10180

Aitchison, J., & Shen, S. M. (1980). Logistic-Normal Distributions: Some Properties and Uses. *Biometrika*, *67*(2), 261–272. https://doi.org/10.1093/biomet/67.2.261

Barker, R. J., Schofield, M. R., Link, W. A., & Sauer, J. R. (2018). On the reliability of N-mixture models for count data. *Biometrics*, *74*(1), 369–377. https://doi.org/10.1111/biom.12734

Betancourt, M. (2018). *A Conceptual Introduction to Hamiltonian Monte Carlo* (arXiv:1701.02434). arXiv. https://doi.org/10.48550/arXiv.1701.02434

Bürkner, P.-C., Gabry, J., & Vehtari, A. (2021). Efficient leave-one-out cross-validation for Bayesian non-factorized normal and Student-t models. *Computational Statistics*, *36*(2), 1243–1261. https://doi.org/10.1007/s00180-020-01045-4

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*, 1. https://doi.org/10.18637/jss.v076.i01

Cormack, R. M. (1964). Estimates of survival from the sighting of marked animals. *Biometrika*, *51*(3/4), 429–438. https://doi.org/10.2307/2334149

de Valpine, P., Turek, D., Paciorek, C., Anderson-Bergman, C., Temple Lang, D., & 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. https://doi.org/10.1080/10618600.2016.1172487

Dorazio, R. M. (2020). Objective prior distributions for Jolly-Seber models of zero-augmented data. *Biometrics*, *76*(4), 1285–1296. https://doi.org/10.1111/biom.13221

Duarte, A., Brown, D. J., & Forstner, M. R. J. (2011). Estimating abundance of the endangered Houston toad on a primary recovery site. *Journal of Fish and Wildlife Management*, *2*(2), 207–215. https://doi.org/10.3996/072011-JFWM-041

Dupuis, J. A., & Schwarz, C. J. (2007). A Bayesian approach to the multistate Jolly-Seber capture-recapture model. *Biometrics*, *63*(4), 1015–1022. https://www.jstor.org/stable/4541454

Ergon, T., Borgan, Ø., Nater, C. R., & Vindenes, Y. (2018). The utility of mortality hazard rates in population analyses. *Methods in Ecology and Evolution*, *9*(10), 2046–2056. https://doi.org/10.1111/2041-210X.13059

Gabry, J., Češnovar, R., Johnson, A., & Bronder, S. (2025). *cmdstanr: R Interface to 'CmdStan'* [Manual].

Gilbert, N. A., Clare, J. D. J., Stenglein, J. L., & Zuckerberg, B. (2021). Abundance estimation of unmarked animals based on camera-trap data. *Conservation Biology*, *35*(1), 88–100.

https://doi.org/10.1111/cobi.13517

Glennie, R., Adam, T., Leos-Barajas, V., Michelot, T., Photopoulou, T., & McClintock, B. T. (2023). Hidden Markov models: Pitfalls and opportunities in ecology. *Methods in Ecology and Evolution*, *14*(1), 43–56. https://doi.org/10.1111/2041-210X.13801

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*, 1593–1623.

Hollanders, M., Grogan, L. F., Nock, C. J., McCallum, H. I., & Newell, D. A. (2023). Recovered frog populations coexist with endemic *Batrachochytrium Dendrobatidis* despite load-dependent mortality. *Ecological Applications*, *33*(1), e2724. https://doi.org/10.1002/eap.2724

Hollanders, M., & Royle, J. A. (2022). Know what you don't know: Embracing state uncertainty in disease-structured multistate models. *Methods in Ecology and Evolution*, *13*(12), 2827–2837. https://doi.org/10.1111/2041-210X.13993

Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration-stochastic model. *Biometrika*, *52*(1/2), 225–247. https://doi.org/10.2307/2333826

Kallioinen, N., Paananen, T., Bürkner, P.-C., & Vehtari, A. (2023). Detecting and diagnosing prior and likelihood sensitivity with power-scaling. *Statistics and Computing*, *34*(57). https://doi.org/10.1007/s11222-023-10366-5

Kellner, K. F., Fowler, N. L., Petroelje, T. R., Kautz, T. M., Beyer, D. E., & Belant, J. L. (2021). ubms: An R package for fitting hierarchical occupancy and N-mixture abundance models in a Bayesian framework. *Methods in Ecology and Evolution*, *13*, 577–584.

Lebreton, J.-D., Burnham, K. P., Clobert, J., & Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs*, *62*(1), 67–118. https://doi.org/10.2307/2937171

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. (2000). WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–337. https://doi.org/10.1023/A:1008929526011

Martin, G. M., Frazier, D. T., & Robert, C. P. (2024). Computing Bayes: From then 'til now'. *Statistical Science*, *39*(1). https://doi.org/10.1214/22-STS876

McClintock, B. T., Langrock, R., Gimenez, O., Cam, E., Borchers, D. L., Glennie, R., & Patterson, T. A. (2020). Uncovering ecological state dynamics with hidden Markov models. *Ecology Letters*, *23*(12), 1878–1903. https://doi.org/10.1111/ele.13610

Michelot, T. (2025). *hmmTMB: Hidden Markov models with flexible covariate effects in R* (arXiv:2211.14139). arXiv. https://doi.org/10.48550/arXiv.2211.14139

Modrák, M., Moon, A. H., Kim, S., Bürkner, P., Huurre, N., Faltejsková, K., Gelman, A., & Vehtari, A. (2023). Simulation-based calibration checking for Bayesian computation: The choice of test quantities shapes sensitivity. *Bayesian Analysis*, *-1*(-1). https://doi.org/10.1214/23-BA1404

Monnahan, C. C., Thorson, J. T., & Branch, T. A. (2017). Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. *Methods in Ecology and Evolution*, *8*(3), 339–348. https://doi.org/10.1111/2041-210X.12681

Neal, R. M. (2011). MCMC using Hamiltonian dynamics. In S. Brooks, A. Gelman, G. Jones, & X.-L. Meng (Eds.), *Handbook of Markov Chain Monte Carlo*. CRC Press. https://doi.org/10.1201/b10905

Ogle, K., & Barber, J. J. (2020). Ensuring identifiability in hierarchical mixed effects Bayesian models. *Ecological Applications*, *30*(7), e02159. https://doi.org/10.1002/eap.2159

Petersen, C. G. J. (1896). The yearly immigration of young plaice in the Limfjord from the German sea. *Report of the Danish Biological Station to the Board of Agriculture*, *6*, 1–48.

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

Pollock, K. H. (1982). A capture-recapture design robust to unequal probability of capture. *The Journal of Wildlife Management*, *46*(3), 752. https://doi.org/10.2307/3808568

Pradel, R. (2005). Multievent: An extension of multistate capture–recapture models to uncertain states. *Biometrics*, *61*(2), 442–447. https://doi.org/10.1111/j.1541-0420.2005.00318.x

R Core Team. (2025). *R: A language and environment for statistical computing* [Manual]. R Foundation for Statistical Computing.

Royle, J. A. (2009). Analysis of capture-recapture models with individual covariates using data augmentation. *Biometrics*, *65*(1), 267–274. https://www.jstor.org/stable/25502266

Royle, J. A., & Dorazio, R. M. (2012). Parameter-expanded data augmentation for Bayesian analysis of capture–recapture models. *Journal of Ornithology*, *152*(2), 521–537. https://doi.org/10.1007/s10336-010-0619-4

Royle, J. A., & Link, W. A. (2006). Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, *87*(4), 835–841. https://doi.org/10.1890/0012-9658(2006)87%5B835:GSOMAF%5D2.0.CO;2

Schwarz, C. J., & Arnason, A. N. (1996). A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics*, *52*(3), 860–873. https://doi.org/10.2307/2533048

Seber, G. A. F. (1965). A Note on the Multiple-Recapture Census. *Biometrika*, *52*(1/2), 249–259. https://doi.org/10.2307/2333827

Socolar, J. B., & Mills, S. C. (2023). *Introducing flocker: An R package for flexible occupancy modeling via brms and Stan*. Ecology. https://doi.org/10.1101/2023.10.26.564080

Socolar, J. B., & Mills, S. C. (2024). *flocker: Flexible occupancy estimation with Stan* [Manual]. https://doi.org/10.32614/CRAN.package.flocker

Stan Development Team. (n.d.). *Stan User's Guide* (2.36 ed.).

Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., & Gelman, A. (2025). *loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models*.

White, G. C., & Burnham, K. P. (1999). Program MARK: Survival estimation from populations of marked animals. *Bird Study*, *46*, S120–S139. https://doi.org/10.1080/00063659909477239

Zucchini, W., MacDonald, I. L., & Langrock, R. (2017). *Hidden Markov Models for Time Series: An Introduction Using R, Second Edition* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/b20790

Box 1: Prior distributions for CR parameters

Bayesian analysis aims to quantify the posterior distribution of model parameters (θ) given the data (y), $\pi(\theta \mid y)$, which is proportional to the product of the prior distribution $\pi(\theta)$ and the likelihood $\pi(y \mid \theta)$. Prior distributions need to be specified for all top-level parameters and should be consistent with our existing beliefs about the system under study. Ideally, these priors are *weakly informative*: broad enough not to dominate the posterior, while still ruling out completely implausible values.

In basic CJS models, mortality rates can be given $\text{Gamma}(1, 1)$ priors, which are uniform on $[0, 1]$ on the probability scale; however, their appropriateness depends on the chosen survey interval. For example, this prior would be implausible for a long-lived species sampled with monthly surveys, as the prior expected annual survival would be $\exp(-1 \cdot 12) \approx 0.0001$. Detection probabilities can be given diffuse Beta priors, where $\text{Beta}(1, 1)$ is uniform on $[0, 1]$. Fixed effect coefficients for mortality, transition, and entry rates (log-link) and detection probabilities (logit-link) can often be given $\text{Normal}(0, \sigma)$ priors, for example with $\sigma = 1$ when covariates are appropriately scaled. Larger values of σ place increasing prior mass on values close to 0 and on very large values (for rates) or close to 1 (for probabilities).

In multistate models, the survey frequency should be chosen so that state transitions are rare relative to the survey intervals. This ensures the *snapshot property*—that the system is effectively static during each survey interval—and we can follow the rule of thumb proposed by Glennie et al. (2023): the mean survey interval should be shorter than the smallest expected dwell time, defined as the reciprocal of the largest transition rate:

$$\bar{\tau} \leq \frac{1}{\max(\boldsymbol{q})}. \tag{17}$$

Therefore, the priors for transition rates $\boldsymbol{q}$ should concentrate most of their mass below $1/\bar{\tau}$, assuming survey intervals are ecologically informed. If $\bar{\tau} = 1$, or when survey intervals are rescaled to have mean 1 before model fitting by dividing by the geometric mean,

$$\boldsymbol{\tau'} = \frac{\boldsymbol{\tau}}{\exp\left(\sum_{j=1}^{J-1} \log \tau_j\right)} \tag{18}$$

then e.g. $\text{Gamma}(1, 3)$ and $\text{Gamma}(2, 5)$ both place approximately 95% of the prior mass in the range satisfying the snapshot property, with the former prior more strongly favouring infrequent transitions. The same scaling simplifies prior choice for JS models, where entry probabilities $\boldsymbol{\beta} \sim \text{Dirichlet}(\boldsymbol{\alpha})$ can be parameterised as $\alpha_1 = \mu\gamma$ and $\alpha_j = \mu\tau_{j-1}, j > 1$. Choosing $\mu \sim \text{Gamma}(1, \bar{\tau})$ and $a = 1$ produces a uniform prior for $\boldsymbol{\beta}$ scaled by the survey intervals on the original scale, while $\text{Gamma}(1, 1)$ is the natural default with scaled $\boldsymbol{\tau'}$. As γ essentially governs the starting population size during $j = 1$ relative to the super-population, a weakly informative choice would be $\text{Gamma}(1, 1)$, as it allows both large and small starting population sizes but places most of the prior density on similar entry probabilities as during the rest of the study.

Temporal variation in model parameters not captured by covariates can be modeled in various ways, with partial pooling using random effects (i.e., log- or logit-normal effects) being a principled choice. This contrasts with the existing maximum likelihood implementations, where the Bayesian analogue would be independent priors for each survey. A useful default prior for temporal parameters is a Gaussian process (GP) prior, for example with exponentiated quadratic kernel where the covariance between two temporal realisations for a parameter is given by

$$K\left(x \mid \sigma, \ell\right)_{jj'} = \sigma^2 \exp\left(-\frac{\left(x_j - x_{j'}\right)^2}{2\ell^2}\right), \tag{2}$$

where x is input, i.e. survey dates, σ is the marginal standard deviation of the random effect (analogous to conventional Gaussian random effects) and ℓ is the length-scale, which captures the degree to which parameters are temporally correlated. Beyond enabling partial pooling across surveys, this prior explicitly models temporal structure in parameters. To capture temporal variation jointly in multiple model parameters (e.g., mortality and transition rates), a multivariate Gaussian process can be implemented—an approach demonstrated in Case Study 2. The Stan language contains many functions that facilitate these priors easily.