Bayesian non-parametric detection heterogeneity in ecological models

Daniel Turek\(^1\) · Claudia Wehrhahn\(^2\) · Olivier Gimenez\(^3\)

Received: 26 September 2020 / Revised: 12 February 2021 / Accepted: 15 February 2021 / Published online: 22 March 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract
Detection heterogeneity is inherent to ecological data, arising from factors such as varied terrain or weather conditions, inconsistent sampling effort, or heterogeneity of individuals themselves. Incorporating additional covariates into a statistical model is one approach for addressing heterogeneity, but there is no guarantee that any set of measurable covariates will adequately address the heterogeneity, and the presence of unmodelled heterogeneity has been shown to produce biases in the resulting inferences. Other approaches for addressing heterogeneity include the use of random effects, or finite mixtures of homogeneous subgroups. Here, we present a non-parametric approach for modeling detection heterogeneity for use in a Bayesian hierarchical framework. We employ a Dirichlet process mixture which allows a flexible number of population subgroups without the need to pre-specify this number of subgroups as in a finite mixture. We describe this non-parametric approach, then consider its use for modeling detection heterogeneity in two common ecological motifs: capture–recapture and occupancy modeling. For each, we consider a homogeneous model, finite mixture models, and the non-parametric approach. We compare these approaches using simulation, and observe the non-parametric approach as the most reliable method for addressing varying degrees of heterogeneity. We also present two real-data examples, and compare the inferences resulting from each modeling approach. Analyses are carried out using the \texttt{nimble} package for \texttt{R}, which provides facilities for Bayesian non-parametric models.

Keywords Bayesian non-parametrics · Capture–recapture · Markov chain Monte Carlo · \texttt{nimble} · Occupancy models · Statistical ecology
1 Introduction

Inferring species distribution and demography are two key questions in ecology (Begon et al. 2006). However, addressing these questions is challenging when studying animals and plants in their natural environments because of the inherent imperfection in the detection process of species or individuals. When ignored, this issue of imperfect detection can lead to biased estimates of species distribution or survival (Gimenez et al. 2008; Guillera-Arroita 2017). To cope with imperfect detection, occupancy models (MacKenzie et al. 2018) and capture–recapture models (McCrea and Morgan 2014) were developed to provide unbiased estimates of species range and individual survival, with numerous applications in all fields of ecology.

As with any statistical model, inferences made from occupancy and capture–recapture analyses rely on assumptions that must be satisfied at least to a reasonable degree. Common to occupancy and capture–recapture models is the assumption of homogeneity of the detection process, which asserts that there is no unmodelled heterogeneity in species and individual detection probabilities. If ignored, heterogeneity in detection leads to flawed inference in occupancy and capture–recapture models (Royle 2006). Detection heterogeneity can exist due to heterogeneous sampling effort, variation in animal abundance or behavior, site characteristics or even on account of varying observer skills. Ideally, covariates could be measured and incorporated in ecological models to account for detection heterogeneity. However, unexplained variation may still remain, or measuring the relevant covariates may simply be impossible.

When unmodelled heterogeneity exists, it can be accommodated using finite mixtures in which discrete latent variables are used to assign sites or individuals to mixture components (i.e., uniform groups) each characterized by group-specific parameters (Royle 2006; Pledger et al. 2010; Louvrier et al. 2018). In simulation studies, finite mixtures were successful in decreasing bias in occupancy and survival probability estimates that was introduced by heterogeneity in the detection process (Pledger 2005; Louvrier et al. 2018). However, from a practical perspective, the issue remains of selecting the number of mixture components for real-life data analyses, which is not straightforward (Cubaynes et al. 2012; Pohle et al. 2017).

Here, we propose a Bayesian non-parametric (BNP) approach to modeling detection heterogeneity in occupancy and capture–recapture models. BNP models provide a flexible approach that relaxes standard modeling assumptions, such as the choice of a specific parametric kernel in density estimation, or here the choice of a fixed number of groups in finite mixtures. Previous uses of a BNP approach in ecological models include the modeling of wildlife migration patterns (Matechou and Caron 2017; Diana et al. 2020), the estimation of population size (Manrique-Vallier 2016; Dorazio et al. 2008) and that of the probability of remaining in or vacating a given area (Ford et al. 2015).

Perhaps the most widely used BNP model is the Dirichlet process mixture (DPM) model (Ferguson 1973, 1974; Lo 1984; Escobar 1994; Escobar and West 1995), which is a mixture model with infinitely many components. DPM models are a suitable fit for addressing the inherent heterogeneity present in ecological models. We consider DPM models of Bernoulli distributions, with the distribution of the Bernoulli detection probability parameters arising from a Dirichlet process. A DPM model can be represented...
in different, yet equivalent manners. Two of them are the Chinese restaurant process (CRP; Blackwell and MacQueen 1973; Pitman 1995, 1996) and the stick-breaking (SB; Sethuraman 1994) representations.

The flexibility of BNP models is usually translated into a hierarchical model, which relies on Markov chain Monte Carlo algorithms to sample from the resulting posterior distribution. The implementation of these algorithms is usually computationally complex and demanding in standard Bayesian software such as WinBUGS or JAGS (e.g. Ohlssen et al. 2007) and often requires writing specific code to implement the BNP model and sampling algorithms (Ford et al. 2015). Recently, the nimble R package introduced non-parametric functionality to address these difficulties, by supporting the use of non-parametric DPM models. Specialized functions, distributions, and samplers are provided for both the CRP and SB representations. Wehrhahn et al. (2018) contains additional details and examples of using these BNP modeling approaches in nimble.

Here, we focus on the CRP representation to address detection heterogeneity in capture–recapture and occupancy models. The article is organized as follows. In the next section, we present homogeneous capture–recapture and occupancy models, finite mixtures and non-parametric models formulated in a Bayesian framework as hierarchical models. In Sect. 3, we give the MCMC sampling scheme used to fit non-parametric models and introduce the nimble and nimbleEcology R packages. Section 4 presents several simulation studies, using capture–recapture and occupancy data, which validate the ability of our approach to capture detection heterogeneity in occupancy and survival probabilities. Section 5 illustrates our method using data from a study of gray wolf (Canis lupus), in trying to estimate occupancy and survival while accounting for detection heterogeneity. The final section provides general conclusions and discusses the potential of our approach.

2 Models

We consider three different approaches to handling heterogeneity. The first approach uses a homogeneous model, which disregards any heterogeneity and considers all individuals and sites to be identical. The second approach uses finite mixtures of population subgroups, where individuals or sites within each subgroup are identical, distinct population subgroups may differ in one or more characteristics, and the number of subgroups is pre-specified. The homogeneous model is a special case of the finite mixture model, where the population contains only a single subgroup.

The third approach uses a non-parametric representation for modeling individual and site heterogeneity, where the number of population subgroups, the assignment of individuals and sites into subgroups, and the characteristics of each subgroup are determined by the data at the time of model fitting. The non-parametric approach does not require pre-specifying a fixed number of population subgroups, but rather, the number of population subgroups itself is an estimable model parameter. The tendency of the BNP model towards using a larger or smaller number of subgroups is controlled by a concentration parameter, for which we consider a range of prior distributions. We now describe the three model formulations in detail, and provide specifications of ecological capture–recapture and occupancy models using each.
2.1 Homogeneous models

In the homogeneous model, we assume that all $N$ observed individuals of the population are identical in their characteristics. There is no heterogeneity between individuals or sites, and thus for any parameter $\theta$ of the population, all individuals or sites share a common value of $\theta$.

2.2 Homogeneous capture–recapture model

We consider a basic ecological capture–recapture model, for binary-valued observation data of $N$ individuals occurring over $T$ observational time periods. We condition on the first observation of each individual occurring in the time period $t = 1$, although it is straightforward to relax this assumption to allow first-detections to occur in other time periods. We parameterize the model in terms of survival probability between time periods $\phi$, and probability of detection conditional on being alive $p$.

In subsequent models we introduce heterogeneity in $p$, but in the homogeneous model all individuals are characterized by the constant probability of detection $p$. We use a state–space formulation of the model, where binary states $x_{i,t}$ give the alive/dead status of individual $i$ at time $t$, and using binary observation data $y_{i,t}$ for individual $i$ at time $t$. The homogeneous capture–recapture model is written:

\[
\begin{align*}
\phi & \sim \text{Uniform}(0, 1) \\
p & \sim \text{Uniform}(0, 1) \\
i = 1, \ldots, N: \\
x_{i,1} & = 1 \\
x_{i,t} & \sim \text{Bernoulli}(\phi \cdot x_{i,t-1}) \text{ for } t = 2, \ldots, T \\
y_{i,t} & \sim \text{Bernoulli}(p \cdot x_{i,t}) \text{ for } t = 1, \ldots, T
\end{align*}
\]

2.3 Homogeneous occupancy model

We consider a homogeneous static occupancy model for a total of $N$ sites, each observed at $T$ distinct sampling occasions. We parameterize the model in terms of the constant site occupancy probability $\psi$, and the probability of detection conditional on site occupancy $p$. In subsequent models we introduce heterogeneity in $p$, but in the homogeneous model, all sites share the probability of detection $p$. We use a state–space formulation of the model, where binary states $z_i$ give the true occupancy status of site $i$, and using binary observation data $y_{i,t}$ for the observation of site $i$ at sampling occasion $t$. The homogeneous occupancy model is written:
\[ \psi \sim \text{Uniform}(0, 1) \]
\[ p \sim \text{Uniform}(0, 1) \]
\[ i = 1, \ldots, N : \]
\[ z_i \sim \text{Bernoulli}(\psi) \]
\[ y_{i,t} \sim \text{Bernoulli}(p \cdot z_i) \text{ for } t = 1, \ldots, T \]

2.4 Finite mixture models

In a finite mixture model specified as having \( K \geq 2 \) distinct population subgroups, each individual or site is considered to be a member of any of the \( K \) subgroups with equal probability \( 1/K \). We introduce a discrete indicator variable for each individual or site, \( g_i \) for individual or site \( i \), where \( g_i \) denotes the “group” of individual or site \( i \). We use an independent discrete uniform prior distributions over the set \( \{1, 2, \ldots, K\} \) for each \( g_i \), and \( g_i = k \) indicates that individual or site \( i \) is a member of population subgroup \( k \).

Furthermore, each of the \( K \) distinct subgroups may differ in one or more demographic characteristics. Again considering the demographic parameter \( \theta \), we introduce \( K \) model parameters \( \theta_1, \theta_2, \ldots, \theta_K \), which are given independent and identical prior distributions. Then, all individuals or sites belonging to population subgroup \( k \) display \( \theta = \theta_k \). To avoid issues of “label exchanging” between groups, and hence a lack of model identifiability, we also impose the constraint that \( \theta_k \leq \theta_{k+1} \), or that the ordered set of parameters \( \{\theta_1, \theta_2, \ldots, \theta_K\} \) is non-decreasing.

2.5 Finite mixture capture–recapture model

We generalize the homogeneous capture–recapture model given in Sect. 2.1 to a \( K \)-group finite mixture model, to allow heterogeneity in the probability of detection. Specifically, we introduce \( K \) new model parameters \( p_1, p_2, \ldots, p_K \), where \( p_k \) represents the probability of detection for individuals in subgroup \( k \). Each \( p_k \) is given an independent Uniform(0, 1) prior distribution, and we impose the constraint on these parameters that \( p_k \leq p_{k'} \) for \( k < k' \) to ensure model identifiability.

Since \( g_i \) gives the subgroup number which contains individual \( i \), we use probability of detection \( p_{g_i} \) for individual \( i \). Putting this together, the \( K \)-group finite mixture capture–recapture model is written as:
\[ \phi \sim \text{Uniform}(0, 1) \]
\[ p_k \sim \text{Uniform}(0, 1) \text{ for } k = 1, \ldots, K \]
\[ p_k \leq p_{k'} \text{ for } k < k' \]
\[ i = 1, \ldots, N : \]
\[ g_i \sim \text{DiscreteUniform}([1, \ldots, K]) \]
\[ x_{i,1} = 1 \]
\[ x_{i,t} \sim \text{Bernoulli}(\phi \cdot x_{i,t-1}) \text{ for } t = 2, \ldots, T \]
\[ y_{i,t} \sim \text{Bernoulli}(p_{g_i} \cdot x_{i,t}) \text{ for } t = 1, \ldots, T \]

### 2.6 Finite mixture occupancy model

We similarly generalize the homogeneous occupancy model from Sect. 2.1 to a \(K\)-group finite mixture model using parameters \(p_1, p_2, \ldots, p_K\), where \(p_k\) represents the probability of detection for sites in subgroup \(k\). We use the same independent Uniform(0, 1) prior distribution for each \(p_k\), and impose the same constraint that \(p_k \leq p_{k'}\) for \(k < k'\) to ensure identifiability. Thus, \(g_i\) indicates the subgroup which contains site \(i\), which therefore has probability of detection \(p_{g_i}\). The \(K\)-group finite mixture occupancy model is written as:

\[ \psi \sim \text{Uniform}(0, 1) \]
\[ p_k \sim \text{Uniform}(0, 1) \text{ for } k = 1, \ldots, K \]
\[ p_k \leq p_{k'} \text{ for } k < k' \]
\[ i = 1, \ldots, N : \]
\[ g_i \sim \text{DiscreteUniform}([1, \ldots, K]) \]
\[ z_i \sim \text{Bernoulli}(\psi) \]
\[ y_{i,t} \sim \text{Bernoulli}(p_{g_i} \cdot z_i) \text{ for } t = 1, \ldots, T \]

### 2.7 Non-parametric models

Using a non-parametric approach, the pre-specification of a fixed number of subgroups is no longer required. Furthermore, our previous assumption that individual or site assignments to subgroups must follow a discrete uniform distribution is also relaxed. In a non-parametric model the number of subgroups in the population is considered unknown and is inferred from the data. Theoretically, there could be an infinite number of population subgroups, although in practice there will never exceed \(N\) subgroups. This means there cannot exist more population subgroups than the total number of observations. But when there are fewer than \(N\) subgroups, individuals or sites can probabilistically move out of an existing group and into a newly created group, with its own distinct probability of detection.

The possibility of the CRP distribution using as many population subgroups as the total number of observations requires the inclusion of \(N\) distinct demographic
parameters $\theta_1, \ldots, \theta_N$ in our hierarchical non-parametric models. These parameters are assumed to be independent and identically distributed, just as in the finite mixture models.

As with the finite mixture model, the subgroup assignment structure is encoded in indicator variables $g_i$, where $g_i = k$ indicates that individual (or site, hereafter) $i$ belongs to population subgroup $k$. Assume that the first $i - 1$ individuals belong to population subgroups that range from 1 to $K$ and consider the following conditional distribution for $g_i$, where $g_{1:(i-1)} = (g_1, \ldots, g_{i-1})$:

$$
\Pr(g_i = k \mid g_{1:(i-1)}, \alpha) = \begin{cases} 
\frac{n_k}{i - 1 + \alpha}, & k = 1, \ldots, K \\
\frac{\alpha}{i - 1 + \alpha}, & k = K + 1
\end{cases}
$$

(1)

where $n_k = n_k(g_{1:(i-1)})$ denotes the number among the first $i - 1$ individuals which exist in subgroup $k$, i.e., $n_k = \sum_{j=1}^{i-1} I(g_j = k)$ where $I(\cdot)$ is the indicator function, $K = K(g_{1:(i-1)})$ denotes the number of distinct subgroups represented among the first $i - 1$ individuals, and $\alpha > 0$ is the concentration parameter. The discrete distribution (1) for the group assignment of individual $i$ (conditional on the group assignments of individuals 1, ..., $i - 1$) implies that each successive individual is assigned to an existing subgroup with probability proportional to the size of each subgroup, and is assigned to a new subgroup with probability proportional to $\alpha$. The product of the successive conditional distributions given by (1) gives rise to the joint distribution of $g_{1:N}$, which is the Chinese restaurant process (CRP) distribution with concentration parameter $\alpha$ (Blackwell and MacQueen 1973; Pitman 1995, 1996). See Li et al. (2019) for more details and interpretations of the CRP prior distribution.

The strictly positive concentration parameter $\alpha$ of the CRP distribution influences the number of subgroups, through its control over the probability that individuals are assigned into new subgroups. The larger the value of $\alpha$, the more likely new subgroups are to be created. Figure 1 illustrates the effect of $\alpha$ on the number of subgroups created from the CRP distribution, when $N = 60$. For $\alpha = 0.1$ generally only one or two subgroups are created, and infrequently three or more. When $\alpha = 0.5$, the number of subgroups created generally falls between one and six. For $\alpha = 1$, we seldom observe only one subgroup, and instead generally have between two and ten groups.

Rather than fixing the CRP concentration parameter $\alpha$ to a specific value, we will assign a prior distribution to $\alpha$. This allows the degree of heterogeneity within the dataset itself to dictate the plausible range for $\alpha$. We consider three different prior distributions for $\alpha$, which are listed in Table 1 along with their respective means and standard deviations. These prior distributions, as ordered in Table 1, provide successively more prior probability to larger values of $\alpha$. As seen in Fig. 1, this will have effect of shifting the prior distribution for the number of subgroups to the right, or increasing the number of population subgroups that we are likely to observe.

The exact prior distribution for the number of subgroups depends not only on $\alpha$ (or the probabilistic prior distribution for $\alpha$), but also on the total number of observations to be grouped, $N$, which was fixed at $N = 60$ for the distributions shown in Fig. 1. Thus, each prior distribution for $\alpha$ corresponds to a particular distribution for the number of
Fig. 1 Prior probability of number of subgroups induced by the CRP($\alpha$) distribution for different values of $\alpha$, when $N = 60$. A color version is available in the electronic version of this article.

Table 1 Prior distributions considered for the CRP concentration parameter

| Prior distribution | Mean   | Standard deviation |
|--------------------|--------|--------------------|
| Gamma(1, 1)        | 1      | 1                  |
| Gamma(1, 0.1)      | 10     | 10                 |
| Gamma(2, 0.1)      | 20     | 14.14              |

Gamma distributions are parameterized as in terms of shape and rate.

population subgroups, for any value of $N$. To give some insight as to the likely number of subgroups to result, for these different $\alpha$ prior distributions and values of $N$, Fig. 2 presents the mean number of subgroups resulting from each prior distribution for a range of $N$ values.

We see, for example, the Gamma(1, 1) prior distribution never suggests a mean number of subgroups in excess of 10, even for values of $N$ as large as 10,000. Bayesian credible intervals (BCIs) are not shown in Fig. 2 on account of their disparate scales, but we note that even when $N = 10,000$, the Gamma(1, 1) prior produces a 95% BCI for the number of subgroups of $(1, 31)$, essentially allocating all prior probability to grouping structures with fewer than 31 subgroups. Also when $N = 10,000$, the 95% BCI resulting from the Gamma(1, 0.1) distribution spans the broad range of $(3, 208)$, while that of the Gamma(2, 0.1) distribution allocates negligible probability to any small number of subgroups, with a 95% BCI of $(19, 297)$. Thus, the selection of the prior distribution for $\alpha$ has an important and profound effect on the population grouping structures which may result, and will be a thoughtful consideration hereafter. The non-parametric hierarchical models presented below denote use of the Gamma(1, 1) prior distribution for $\alpha$, but in all applications we consider the full range of prior distributions listed in Table 1.
Fig. 2 Mean number of population subgroups resulting from each prior distribution for the CRP concentration parameter $\alpha$, for different total numbers of observations $N$.

### 2.8 Non-parametric capture–recapture model

In the non-parametric capture–recapture model, detection heterogeneity is flexibly addressed using a CRP prior distribution to allow for an unknown number of population subgroups. The number of population subgroups will thereby be inferred from the data, and is not fixed to a pre-specified value as in the finite mixture models described in Sect. 2.4. We assign a CRP($\alpha$) prior distribution to $g_{1:N}$, the vector of subgroup indicator variables, and include a prior distribution for $\alpha$.

We introduce probabilities of detection $p_1, p_2, \ldots, p_N$, each independent and identically distributed, following a Uniform(0, 1) prior distribution. In general, fewer than $N$ subgroups are actually used by the CRP distribution, and thus only a subset of the probabilities of detection are “active” in terms of their influence on the model likelihood calculation. As in the finite mixture models, $g_i = k$ indicates that individual $i$ is a member of subgroup $k$, and therefore has probability of detection $p_{g_i}$. The full non-parametric capture–recapture model is written as:

\[
\begin{align*}
\phi & \sim \text{Uniform}(0, 1) \\
p_k & \sim \text{Uniform}(0, 1) \quad \text{for } k = 1, \ldots, N \\
\alpha & \sim \text{Gamma}(1, 1) \\
g_{1:N} & \sim \text{CRP}(\alpha) \\
i = 1, \ldots, N: \\
x_{i,1} & = 1 \\
x_{i,t} & \sim \text{Bernoulli}(\phi \cdot x_{i,t-1}) \quad \text{for } t = 2, \ldots, T \\
y_{i,t} & \sim \text{Bernoulli}(p_{g_i} \cdot x_{i,t}) \quad \text{for } t = 1, \ldots, T
\end{align*}
\]
2.9 Non-parametric occupancy model

Similarly, we generalize the finite mixture occupancy model to use a CRP($\alpha$) prior distribution for individual group assignments $g_{1:N}$, and include a prior distribution for $\alpha$. We include the maximum possible necessary number of distinct probabilities of detection $p_1, p_2, \ldots, p_N$, each with independent Uniform(0, 1) prior distributions. Here, $g_i = k$ indicates that site $i$ is a member of subgroup $k$, and therefore has probability of detection $p_{g_i}$. The full non-parametric occupancy model is written as:

$$\begin{align*}
\psi & \sim \text{Uniform}(0, 1) \\
p_k & \sim \text{Uniform}(0, 1) \text{ for } k = 1, \ldots, N \\
\alpha & \sim \text{Gamma}(1, 1) \\
g_{1:N} & \sim \text{CRP}(\alpha) \\
i & = 1, \ldots, N: \\
z_i & \sim \text{Bernoulli}(\psi) \\
y_{i,t} & \sim \text{Bernoulli}(p_{g_i} \cdot z_i) \text{ for } t = 1, \ldots, T
\end{align*}$$

3 Model fitting via Markov chain Monte Carlo

Mathematical descriptions of our approaches to modeling heterogeneity within a population were given in Sect. 2. These models are formulated in a Bayesian framework as hierarchical models. Prior distributions are specified for model parameters, and the conditional distributions for discrete latent states (i.e., dead/alive status) and the data likelihoods are given. The general tool for fitting hierarchical models to data is Markov chain Monte Carlo (MCMC; Brooks et al. 2011), a stochastic sampling algorithm for generating posterior samples for model parameters, conditional on the data. Demographic inferences are subsequently performed using the empirical posterior distributions of model parameters.

Markov chain Monte Carlo is a powerful tool for fitting generally requires the use of software. We make use of the recently developed nimble R package (NIMBLE Development Team 2019), which offers new degrees of freedom for algorithmic development and customization of MCMC sampling (de Valpine et al. 2017). The ability to write custom distributions for use in hierarchical models, and the flexible nature of nimble’s MCMC engine has provided noteworthy gains in fitting ecological models (Turek et al. 2016), and more generally in the study of MCMC algorithms (Turek et al. 2017).

We make use of nimble’s MCMC engine for fitting these hierarchical models, after expressing them in the BUGS language (Lunn et al. 2009). In addition, we make use of two aspects of nimble’s flexibility. For modeling of individual heterogeneity, we make use of the CRP distribution and corresponding sampling algorithms, which are recent additions to the nimble package. Second, we increase sampling efficiency of the specific ecological models by using custom likelihood distributions provided by
the nimbleEcology package (Goldstein et al. 2020) to remove latent states from the model structures.

3.1 Non-parametric distributions and MCMC sampling

When a hierarchical model is formulated using the CRP distribution, the dCRP distribution (available in the nimble package) assigns the joint prior distribution arising from (1) to the labeling vector, $g$. Correspondingly, a specialized sampler is assigned by nimble's MCMC engine. Because the likelihood function of the ecological models presented in Sect. 2 are not conjugate for the prior distribution of the cluster parameters, $p_k$, the non-conjugate sampling algorithm described in Algorithm 8 of Neal (2000) is assigned to $g$.

As described in Sect. 2, under a non-parametric approach, the number of population subgroups is not fixed. In terms of the MCMC sampling scheme this means that the number of subgroups, and therefore the number of cluster parameters which are active, can vary with every MCMC iteration. As nimble does not support dynamic length allocation, the number of cluster parameters defined in the model must be fixed. A safe option would be to consider $N$ cluster parameters, however this is highly inefficient both in terms of computation and storage, especially for large values of $N$. To reduce this inefficiency nimble allows the specification of $N' < N$ cluster parameters. If upon any MCMC iteration more than $N'$ groups are created, then a warning is issued. Additionally, to reduce the computational burden of the non-parametric sampling, only the active cluster parameters, $p_k$, are updated.

As discussed in Sect. 2, the concentration parameter $\alpha$ has important implications in the clustering structure of the model. Therefore, efficient sampling of $\alpha$ is an important matter. Although its posterior distribution does not belong to any known class of distributions, when a Gamma prior distribution is assigned to $\alpha$, a computationally efficient sampling scheme (Escobar and West 1995, Sect. 6) is assigned by nimble’s MCMC engine.

3.2 Likelihood distributions using nimbleEcology

To reduce computation time, we make use of the nimbleEcology R package (Goldstein et al. 2020) for specifying the ecological hierarchical models. The nimbleEcology package provides likelihood distributions specific for a variety of common ecological models, which are implemented as custom distributions using the nimble package. The likelihood distributions provided in nimbleEcology include those for capture–recapture models, occupancy and dynamic occupancy models, and more generally for discrete hidden Markov models (HMMs) as appear in multi-state or multi-event capture–recapture (i.e., Turek et al. 2016).

For each type of ecological model, the likelihood distribution provided by nimbleEcology marginalizes over discrete latent states to directly calculate the unconditional likelihood of observed data. This allows removal of discrete latent state variables—$x_{i,t}$ alive/dead indicator variables in capture–recapture, and $z_i$ occupancy indicator variables in occupancy modeling—from the hierarchical model. This reduces
model size and the necessary model computations, and increases the speed of MCMC mixing of top-level model parameters ($\phi$ in capture–recapture, and $\psi$ in occupancy models) to generate stronger posterior inferences in less computational time.

Using the distributions provided in nimbleEcology to remove latent states does not alter the posterior results generated from each model; it only serves to increase the speed of generating robust inferences. The only noteworthy difference is that posterior inference for the discrete latent states cannot be performed, since samples for these latent states are never generated. So, for example, we could not perform inference for the alive/dead status $x_{i,t}$ of specific individuals in the capture–recapture setting.

4 Simulations

We undertake a variety of simulation studies to assess performance of the various approaches to modeling individual and site heterogeneity. Broadly, we consider the context of capture–recapture models, and also that of occupancy models. For each ecological context, we first consider varying degrees of heterogeneity in detection probability between two population subgroups. Second, we consider a more diverse population consisting of four distinct subgroups, each displaying a unique detection probability.

For each of these scenarios, we fit the homogeneous model, finite mixture models with varying numbers of groups, and non-parametric models as were presented in Sect. 2. Our goal is to study the effect of using the various models presented for modeling heterogeneity in detection probability on the accuracy of the resulting inferences.

R code for all simulations, including the nimble specifications for each model using the likelihood distributions provided in the nimbleEcology package, are available on GitHub ([https://github.com/danielturek/bnp-ecology-examples](https://github.com/danielturek/bnp-ecology-examples)). Next we describe the details of each simulation, and then present results.

4.1 Capture–recapture simulations

In the first capture–recapture simulation we consider two distinct population subgroups of equal size, which differ in their probability of detection. We consider total population sizes of 1600 and 200 individuals. We use eight observational periods, and condition on all individuals being sighted on the first observational period. Survival probability is fixed at $\phi = 0.7$ for simulating data, and we focus on the ability of various modeling approaches to estimate survival. One population subgroup is fixed as having individual probability of detection $p_0 = 0.8$, where detection is conditional on being alive. Individuals in the other subgroup have a fixed detection probability $p$, which varies between simulations. We consider values of $p$ between 0.1 and 0.8 in increments of 0.1, where the terminal case $p = 0.8$ coincides with the detection probability of the first subgroup, and therefore represents a homogeneous population.

In the second capture–recapture simulation, we consider four population subgroups each consisting of 50 individuals, for a total population size of 200 individuals. The four population subgroups have individual probabilities of detection $p_1 = 0.2$, $p_2 = 0.4$, and $p_3 = 0.6$, respectively.
\( p_3 = 0.6, \) and \( p_4 = 0.8. \) We again use eight observational periods, condition on all individuals being sighted on the first observational period, and fix survival probability at \( \phi = 0.7. \)

### 4.2 Occupancy simulations

In the first occupancy simulation, we consider two distinct subgroups of sites of equal size, which differ in their probability of detection. We consider two cases of 4,000 or 400 sites in total. We use six independent observations of each site. The proportion of occupied sites is fixed at \( \psi = 0.7 \) for simulating data, and we focus on the ability of various modeling approaches to estimate the true occupancy proportion. One subgroup is fixed as having probability of detection \( p_0 = 0.8 \) on each independent site visit, where detection is conditional on a site being occupied. The other subgroup has a fixed detection probability \( p \), which varies between simulations. We consider values of \( p \) between 0.1 and 0.8 in increments of 0.1, but with a finer resolution on the interval between 0.1 and 0.4. Again, the terminal case \( p = 0.8 \) coincides with the detection probability of the first subgroup, and therefore represents a homogeneous population.

In the second occupancy simulation, we consider four population subgroups each consisting of 100 sites, for a total of 400 sites. The four population subgroups have individual site probabilities of detection \( p_1 = 0.2, p_2 = 0.4, p_3 = 0.6, \) and \( p_4 = 0.8. \) We again use six independent observations of each site, and fix the proportion of occupied sites at \( \psi = 0.7. \)

### 4.3 Simulation results

Posterior inferences were performed for individual survival probability \( \phi \) in the capture–recapture simulation and for site occupancy proportion \( \psi \) in the occupancy simulation. We use the posterior median and equal-tailed 95% Bayesian credible intervals for inferences under each model. Fitted models include the homogeneous model, finite mixture models with varying numbers of subgroups, and Bayesian non-parametric models using the prior distributions listed in Table 1. Each model was fit to simulated data using MCMC, as described in Sect. 3.

For the simulations which include four distinct population subgroups, in addition to parameter inferences under each model, we also present the WAIC value (Watanabe 2010; Gelman et al. 2014) of each fitted model. WAIC is a measure of the goodness-of-fit of a hierarchical model structure to a given dataset, calculated using chains of posterior MCMC samples. WAIC is measured on the scale of deviance, and therefore lower values of WAIC indicate a more parsimonious fit to the data. For ease of interpretation, we subtract the smallest WAIC value from the WAIC of each model, such that the most parsimonious model displays a WAIC value of zero.

#### 4.3.1 Capture–recapture simulation results

Results for the capture–recapture simulation with two population subgroups and a total of 1,600 individuals appear in Fig. 3. The homogeneous model consistently under-
estimates $\phi$, more severely for larger degrees of heterogeneity between the two groups, with the discrepancy diminishing and disappearing as the detection probabilities of the two groups converge. For the lowest value of $p = 0.1$, and hence a large difference in detection probability between groups, the 3-group mixture model has a slight tendency to inflate estimates of $\phi$, while the opposite is true of the 2-group mixture model; however, this difference is minor and disappears as the group difference decreases. Otherwise, the mixture models and non-parametric model are all comparably successful in generating accurate inferences of $\phi$. The non-parametric models using the other two prior distributions were qualitatively identical to the non-parametric model using the Gamma(1, 1) prior distribution, and are therefore not shown. We also note that all models exhibit a slight negative bias in their estimates of $\phi$. This results from the imperfect detection process on the final observation period, which confers a small negative bias in the estimation of survival. This bias would vanish when detection probability approaches one, or when the number of observation periods diverges.

Results for the capture–recapture simulation with two population subgroups and total population size of 200 individuals appear in Fig. 4. We observe a slightly noisier version of qualitatively similar results to the case of the larger population size, in which inferences were more refined and gave a more precise picture of each model’s ability to address heterogeneity. Once again, the homogeneous model under-estimates $\phi$, more severely for greater degrees of heterogeneity, and the non-parametric and finite mixture models perform similarly well. The non-parametric models using the other two Gamma prior distributions again performed similarly to the non-parametric model shown.

Results for the capture–recapture simulation with a more heterogeneous population, with four equally-sized subgroups among a population of 200 individuals, appear in
Fig. 4 Capture–recapture simulation results for survival probability $\phi$ using two simulated population subgroups and $N = 200$ individuals. One subgroup has probability of detection $p_0 = 0.8$, and the other has probability of detection $p$. Solid lines show posterior median estimates of $\phi$, and dashed lines show upper and lower limits of a 95% Bayesian credible interval. A color version is available in the electronic version of this article.

Table 2. The homogeneous model exhibits the highest WAIC value (indicating the worst fit to the data), and under-estimates survival probability. Otherwise, all models performed nearly identically in their ability to estimate survival. All models were fitted to the same simulated dataset, so this particular dataset appears to most strongly support the value $\phi = 0.71$. The non-parametric model using the Gamma(2, 0.1) prior distribution produced the lowest WAIC value, with the 10-group mixture model yielding the second-lowest WAIC. These are closely followed by the 9-group mixture model and the non-parametric model using the Gamma(1, 0.1) prior distribution. Noting that the Gamma(1, 0.1) and the Gamma(2, 0.1) prior distributions provide the most support to larger numbers of subgroups, it appears the data reasonably support the existence of ten or more population subgroups. Rather than continuing to fit finite mixture models with even larger numbers of subgroups, it would be safe and reasonable to base inferences on any of the non-parametric models.

4.3.2 Occupancy simulation results

Results for the occupancy simulation with two population subgroups and a total of 4000 sites appear in Fig. 5, which only displays the non-parametric model using the Gamma(1, 1) prior distribution. Once again, for low values of $p$, the homogeneous model under-estimates $\psi$, more severely for larger degrees of heterogeneity. The 2-group mixture model (which is also the data-generating model) generates reliable inferences for all values of $p$.

We also see a regime of $p$ values approximately between 0.167 and 0.267, in which the 3-group mixture model vastly over-estimates $\psi$, with posterior median estimates in excess of 0.99. The location of this regime varied somewhat depending on simulation...
Table 2. Capture–recapture simulation results for survival probability $\phi$ using four simulated population subgroups and 200 individuals.

| Model                        | Median | 95% BCI   | WAIC |
|------------------------------|--------|-----------|------|
| Homogeneous                  | .68    | (.63, .72)| 27.4 |
| Finite Mixture ($K = 2$)     | .70    | (.65, .76)| 7.1  |
| Finite Mixture ($K = 3$)     | .71    | (.65, .76)| 6.2  |
| Finite Mixture ($K = 4$)     | .71    | (.66, .76)| 5.6  |
| Finite Mixture ($K = 5$)     | .71    | (.66, .76)| 4.4  |
| Finite Mixture ($K = 6$)     | .71    | (.66, .76)| 3.2  |
| Finite Mixture ($K = 7$)     | .71    | (.66, .76)| 2.3  |
| Finite Mixture ($K = 8$)     | .71    | (.66, .76)| 1.9  |
| Finite Mixture ($K = 9$)     | .71    | (.66, .76)| 1.3  |
| Finite Mixture ($K = 10$)    | .71    | (.66, .76)| 0.8  |
| BNP Gamma(1, 1)              | .71    | (.66, .76)| 7.8  |
| BNP Gamma(1, 0.1)            | .71    | (.66, .76)| 1.5  |
| BNP Gamma(2, 0.1)            | .71    | (.66, .76)| 0.0  |

Subgroups have individual probabilities of detection $p_1 = 0.2$, $p_2 = 0.4$, $p_3 = 0.6$, and $p_4 = 0.8$. Posterior median and 95% Bayesian credible intervals are provided, as well as the WAIC value of each fitted model.

Fig. 5. Occupancy simulation results for site occupancy $\psi$ using two simulated population subgroups and $N = 4,000$ individuals. One subgroup has probability of detection $p_0 = 0.8$, and the other has probability of detection $p$. Solid lines show posterior median estimates of $\psi$, and dashed lines show upper and lower limits of a 95% Bayesian credible interval. A color version is available in the electronic version of this article.

Parameters, but its existence persisted in all simulations, and thus this result is not due to a particular anomalous simulated dataset. The posterior results suggest this exists because the 3-group mixture model admits a third subgroup with effectively zero detection probability, which consumes all structurally unoccupied and no-detection sites, and the non-trivial membership of this group suggests all sites being occupied.
Fig. 6  Occupancy simulation results for site occupancy ψ using two simulated population subgroups and $N = 4,000$ individuals. One subgroup has probability of detection $p_0 = 0.8$, and the other has probability of detection $p$. Solid lines show posterior median estimates of ψ, and dashed lines show upper and lower limits of a 95% Bayesian credible interval. A color version is available in the electronic version of this article (ψ ≈ 1). However, for larger values of $p$, few (if any) occupied sites have zero positive detections, reducing support for this grouping structure. Importantly, however, we see there exists the potential for highly inaccurate inferences when the number of groups in a finite mixture model is chosen either too high, or too low.

While Fig. 5 included only one non-parametric prior distribution, results from all three prior distributions are shown in Fig. 6. The vertical scale is consistent with Fig. 5 for ease of comparison. We see that the upper 95% confidence limit of all three models also fluctuates inside the same regime of low $p$ values, as the non-parametric models consider groupings similar to the 3-group mixture model. However, importantly, the posterior median of all non-parametric models remains reasonable always, and these models provide sensible inferences. We also note that the prior distributions which provide more support for larger numbers of groups display more fluctuation in the upper confidence limit, while use of the Gamma(1, 1) prior distribution somewhat mitigates this effect. If one did not know the true population structure, then the non-parametric model using the Gamma(1, 1) prior distribution appears to be a safe and reliable choice.

Results for the occupancy simulation with two population subgroups and a total of 400 sites appear in Fig. 7. Again, these results exhibit more noise than when using a larger number of simulated sites. The homogeneous model consistently underestimates ψ, in particular for larger degrees of heterogeneity, and the 3-group mixture model still displays a regime of inflated estimates for ψ in the situation of highly disparate subgroups. The upper limit of the non-parametric model is also affected by this, however the non-parametric posterior median remains reasonable always, similar to that of the 2-group mixture model. Results from the non-parametric models using
Fig. 7 Occupancy simulation results for site occupancy $\psi$ using two simulated population subgroups and $N = 400$ individuals. One subgroup has probability of detection $p_0 = 0.8$, and the other has probability of detection $p$. Solid lines show posterior median estimates of $\psi$, and dashed lines show upper and lower limits of a 95% Bayesian credible interval. A color version is available in the electronic version of this article

the other two prior distributions were again extremely similar to the non-parametric model shown, and are not displayed.

Results for the occupancy simulation using four equally-sized population subgroups among 400 total sites appear in Table 3. The homogeneous model exhibits the highest WAIC value by far, and is not considered as a viable candidate. We find the strongest WAIC support for the finite mixture models of $K \geq 7$ subgroups and for the non-parametric models which favor larger numbers of subgroups, suggesting the data support a high degree of detection heterogeneity within the population. All of these models give similar inferences for $\psi$, also noting that this particular dataset appears to reflect site occupancy of $\phi \approx 0.73$, slightly larger than the value used for data simulation. The data-generating model, the 4-group mixture model, has a slightly larger WAIC value but also produces comparable posterior results. Any of these models would be a sensible choice to use for inference.

5 Examples

We present two real-data examples, one in the context of capture–recapture and the other in that of occupancy modeling. For each example, we fit a homogeneous model for detection probability, finite mixture models containing between two and ten population subgroups, and non-parametric models considering the three prior distributions listed in Table 1. In addition to parameter inferences under each model, we also include the WAIC value of each fitted model. Finally, for each example, we also present the posterior distribution of the number of population subgroups resulting from the non-parametric model using each prior distribution. These distri-
Table 3  Occupancy simulation results for site occupancy $\psi$ using four simulated population subgroups and 400 sites

| Model                      | Median | 95% BCI       | WAIC |
|----------------------------|--------|---------------|------|
| Homogeneous                | .65    | (.60, .69)    | 159.0|
| Finite Mixture ($K = 2$)   | .68    | (.63, .74)    | 28.3 |
| Finite Mixture ($K = 3$)   | .70    | (.63, .93)    | 13.6 |
| Finite Mixture ($K = 4$)   | .72    | (.64, .90)    | 9.3  |
| Finite Mixture ($K = 5$)   | .71    | (.64, .90)    | 11.4 |
| Finite Mixture ($K = 6$)   | .72    | (.64, .89)    | 6.5  |
| Finite Mixture ($K = 7$)   | .73    | (.65, .93)    | 3.1  |
| Finite Mixture ($K = 8$)   | .74    | (.65, .88)    | 3.1  |
| Finite Mixture ($K = 9$)   | .75    | (.65, .90)    | 0.0  |
| Finite Mixture ($K = 10$)  | .73    | (.65, .87)    | 3.8  |
| BNP Gamma(1, 1)            | .71    | (.64, .89)    | 18.3 |
| BNP Gamma(1, 0.1)          | .73    | (.65, .87)    | 8.9  |
| BNP Gamma(2, 0.1)          | .73    | (.66, .87)    | 5.0  |

Subgroups have individual probabilities of detection $p_1 = 0.2$, $p_2 = 0.4$, $p_3 = 0.6$, and $p_4 = 0.8$. Posterior median and 95% Bayesian credible intervals are provided, as well as the WAIC value of each fitted model.

Distributions shed light on the degree of heterogeneity present in the data, as inferred by the non-parametric model, and how this distribution differs depending on the choice of prior distribution. The datasets used for each example are available on GitHub at https://github.com/danielturek/bnp-ecology-examples.

5.1 Capture–recapture example

We consider wolf (Canis lupus) capture–recapture data collected in France between 1995 and 2003, as studied in Cubaynes et al. (2010). The original data contains binary detection data for a total of 87 wolves, over $T = 8$ observation periods. Since we condition on the first sighting of each individual, we excluded the twenty individuals who were first observed on the final observation period. This leaves a total of $N = 67$ unique individuals in the dataset. Models for detection heterogeneity described in Sect. 2 were fit to this data using MCMC, and inference was performed for individual survival probability $\phi$ under each model.

Posterior median and 95% credible intervals, as well as the WAIC value of each fitted model are presented in Table 4. The homogeneous model produces the largest WAIC value among those models considered (indicating the poorest fit to the data), and the lowest estimates of $\phi$. Posterior inferences from all mixture models and the non-parametric models are nearly indistinguishable, with posterior median values around 0.91, and 95% credible intervals of approximately (0.79, 0.99).

We note that the non-parametric model using the Gamma(1, 1) prior distribution produces the third-lowest WAIC value, after the 4-group and 5-group mixture models. Furthermore, the inferences for $\phi$ generated using this non-parametric model are practically identical to those of the 4-group and 5-group mixture models. This Gamma(1, 1)
Table 4  Capture–recapture example results

| Model                      | Median | 95% BCI     | WAIC |
|----------------------------|--------|-------------|------|
| Homogeneous                | .80    | (.70, .89)  | 39.2 |
| Finite Mixture (K = 2)     | .91    | (.79, .99)  | 10.8 |
| Finite Mixture (K = 3)     | .91    | (.80, .99)  | 3.6  |
| Finite Mixture (K = 4)     | .92    | (.80, .99)  | 0.0  |
| Finite Mixture (K = 5)     | .92    | (.80, .99)  | 0.0  |
| Finite Mixture (K = 6)     | .91    | (.80, .99)  | 1.6  |
| Finite Mixture (K = 7)     | .91    | (.79, .99)  | 2.3  |
| Finite Mixture (K = 8)     | .90    | (.79, .98)  | 3.2  |
| Finite Mixture (K = 9)     | .90    | (.79, .98)  | 3.8  |
| Finite Mixture (K = 10)    | .90    | (.78, .98)  | 4.5  |
| BNP Gamma(1, 1)            | .92    | (.81, .99)  | 1.5  |
| BNP Gamma(1, 0.1)          | .92    | (.80, .99)  | 2.9  |
| BNP Gamma(2, 0.1)          | .91    | (.79, .99)  | 4.5  |

Posterior inferences are for individual survival probability \( \phi \), and WAIC values indicate the goodness-of-fit of each model.

Prior distribution is the most conservative prior in terms of the number of population subgroups, assigning more prior support to fewer groups. Given our uncertainty in the structure and degree of heterogeneity, this non-parametric model provides defensible inferences and goodness-of-fit. The non-parametric models which assign more prior support to larger numbers of population subgroups produced WAIC values similar to the mixture models containing more groups, as we might expect.

Figure 8 displays the posterior distributions for the number of population subgroups, as inferred by the non-parametric model using each \( \alpha \) prior distribution. Indeed, using of the Gamma(1, 1) prior results in most typically between four and eight subgroups, with a mode of six, and rarely more than fifteen. We also observe the strong tendency of the other two prior distributions to produce significantly larger numbers of population subgroups, with the Gamma(1, 0.1) prior often producing more than 20,
Table 5 Occupancy model example results

| Model                  | Median  | 95% BCI       | WAIC |
|------------------------|---------|---------------|------|
| Homogeneous            | .063    | (.054, .073)  | 70.3 |
| Finite Mixture (K = 2) | .079    | (.063, .101)  | 29.1 |
| Finite Mixture (K = 3) | .087    | (.067, .122)  | 11.6 |
| Finite Mixture (K = 4) | .092    | (.068, .139)  | 2.0  |
| Finite Mixture (K = 5) | .089    | (.068, .132)  | 5.7  |
| Finite Mixture (K = 6) | .090    | (.069, .130)  | 5.4  |
| Finite Mixture (K = 7) | .089    | (.068, .125)  | 7.0  |
| Finite Mixture (K = 8) | .087    | (.068, .122)  | 8.2  |
| Finite Mixture (K = 9) | .087    | (.068, .119)  | 9.2  |
| BNP Gamma(1, 1)        | .090    | (.066, .359)  | 0.0  |
| BNP Gamma(1, 0.1)      | .091    | (.067, .192)  | 0.3  |
| BNP Gamma(2, 0.1)      | .090    | (.068, .154)  | 3.5  |

Posterior inferences are for site occupancy proportion \( \psi \), and WAIC values indicate the goodness-of-fit of each model.

and the Gamma(2, 0.1) prior sometimes producing in excess of 30 subgroups. Relating these results back to the prior distributions for the number of subgroups which result from each Gamma prior distribution for \( \alpha \) (Fig. 2), we note that for \( N = 67 \), the Gamma(1, 1) prior corresponds to a prior 95\% BCI for the number of subgroups of (1, 12), the Gamma(1, 0.1) prior corresponds to a prior 95\% BCI of (1, 40), and the the Gamma(2, 0.1) prior corresponds to a prior 95\% BCI of (8, 46).

When fitting a model to \( N = 67 \) total observations, a model which includes 30 or more distinct subgroups may feel excessive, as each group would contain about two individuals on average. For this small number of observations, and considering the resulting WAIC values, use of the non-parametric model with the Gamma(1, 1) prior distribution may be prudent.

5.2 Occupancy example

For the occupancy example, we consider a second wolf (\textit{Canis lupus}) dataset (Louvrier et al. 2018) collected in France in 2013. Opportunistic observational data such as tracks, scat, and prey remains were collected from \( N = 3,211 \) grid cells, each being a 10km \( \times \) 10km square. Each site was surveyed on a total of \( T = 4 \) independent observation occasions. The categorical data, in total, consisted of 250 “unambiguous detections”, 54 “ambiguous detections”, and 12,540 “non-detections”. We convert this to binary data, wherein both “unambiguous” and “ambiguous” detections are considered to be positive detections. Models for detection heterogeneity described in Sect. 2 were fit to this data using MCMC, and inference was performed for site occupancy proportion \( \psi \) under each model.
Posterior median and 95% credible intervals, as well as the WAIC value of each fitted model are presented in Table 5. The homogeneous model again produces the largest WAIC value, and the lowest estimates of $\psi$. The 2-group mixture model yields the second-highest WAIC value and also lower estimates of $\psi$ than the remaining models, and the 3-group mixture model yields the third-highest WAIC value. This suggests a nontrivial degree of heterogeneity in detection probability between sites, as we might expect from a large number of sites. The remaining mixture models ($K \geq 4$ groups) give WAIC values between 2.0 and 9.2, and exhibit only small variations in the inferences for $\psi$.

The non-parametric model using the Gamma(1, 1) prior distribution yields the lowest WAIC value, however the Gamma(1, 0.1) prior produces WAIC only 0.3 higher. Furthermore, the Gamma(1, 1) prior yields a significantly higher upper-bound for the 95% credible interval than any other model considered. We may question this inference, since the Gamma(1, 1) prior is by far the most restrictive in terms of the number of population subgroups, which may not be an accurate prior belief when modeling $N = 3,211$ distinct sites. In contrast, the other two prior distributions yield only slightly wider credible intervals than the finite mixture models, and competitive WAIC values. Considering the large number of sites in this dataset, use of a non-parametric model with a more liberal prior distribution, which assigns more prior support to larger numbers of subgroups, would seem reasonable.

Figure 9 presents the posterior distribution for the number of population subgroups as inferred by the non-parametric model using each prior distribution. We clearly observe the restrictive effect of the Gamma(1, 1) prior distribution, which effectively limits the number of population subgroups to 20 or fewer. In contrast, the other two prior distributions allow the non-parametric model to consider significantly more population subgroups. The most liberal prior, the Gamma(2, 0.1) distribution, results in a posterior median number of groups around 30, and a 95% credible interval, suggesting between ten and eighty distinct subgroups. These numbers are not unreasonable when modeling such a large number of sites. Either prior other than the restrictive Gamma(1, 1) distribution would seem to be a reasonable choice.
6 Discussion

Here, we have employed a Bayesian non-parametric approach to modeling heterogeneity in the detection process of ecological models. Our approach used a Chinese restaurant process (CRP) prior distribution to model group memberships, which has the benefit of not requiring specification of the number of population sub-groups \( a \) priori. Using the CRP prior, both the distribution of individuals or sites among distinct subgroups and the number of subgroups are inferred from the data. This strategy obviates any process of model selection used to choose a “best value” for the number of groups \( K \) when using a finite mixture model.

The tendency of the CRP distribution towards using fewer (and hence more diverse) subgroups, or towards using a larger number of tightly specified subgroups is governed by a concentration parameter \( \alpha \). The effect of different values of \( \alpha \) is seen in Fig. 1, although we avoid the need to choose a particular value of \( \alpha \) by assigning a prior distribution to the concentration parameter itself. This approach of using a CRP prior distribution to model heterogeneity is used in diverse areas of research, including topic modeling (Blei et al. 2010), genomics (Qin 2006), and evolutionary clustering (Ahmed and Xing 2008), among others.

When employing a Bayesian non-parametric model using the CRP distribution, selection of a prior distribution for the CRP concentration parameter is indeed a subjective choice that warrants thought and attention. Herein, we have considered a range of prior distributions for \( \alpha \), and observed the effects of this choice. A prior favoring fewer groups can severely restrict a model’s ability to partition observations into a larger number of subgroups, which may be unreasonable in the case of many observations or prior belief in a high degree of heterogeneity within a population. In contrast, when modeling a modest number of observations, prior distributions which readily consider large numbers of groups can result in unrealistically many groupings of the observations. In practice, we suggest informing this choice by considering the statistical properties of the candidate prior distributions, the sheer number of observations being modeled, and also by fitting models using different prior distributions to assess the sensitivity of the resulting inferences to this choice.

The examples and simulations used herein focus on the most basic forms of capture-recapture and occupancy models. Specifically, we have considered the Cormack–Jolly–Seber capture–recapture model, and a single-season static occupancy model. Although our examples focused on the basic forms of each model, the same techniques are readily applied to more complex variations of each model. For example, we could apply the same non-parametric CRP prior distribution in multi-state capture–recapture, or in dynamic, multi-season, or multi-species occupancy models. Similarly, one can readily extend the hierarchical models to incorporate individual or environmental covariates affecting population demographics, survey effort, or the detection processes, and apply this same technique to a wide range of ecological models where population heterogeneity is a consideration.

That said, non-parametric modeling of detection heterogeneity can lead to problems of non-identifyability for models specified in terms of an unknown population size, where abundance itself is a target of estimation (Link 2003, 2004). In this situation, the subgroup assignments and detection probabilities of unobserved individuals are nearly
unconstrained, which can lead to diverse inferences of abundance. In the capture-recapture models presented here, we avoid this issue by conditioning on the first capture of individuals, and not attempting to infer total population size. The issue of estimating abundance in the presence of detection heterogeneity is addressed elsewhere (e.g., Cubaynes et al. 2010; Warren et al. 2013), and is not the focus of our work. Despite this, we believe the non-parametric approaches presented herein are powerful and general tools for use in ecological models.

We have used the nimble R package to specify the hierarchical models described herein, and to fit these models to data using MCMC. nimble provides degrees of flexibility which are not available in other software packages. Specifically, we leverage nimble’s ability to use custom-written likelihood distributions in a hierarchical model, specifically the ecological likelihood distributions provided in the nimbleEcology package. Further, nimble provides the ability to specify the sampling algorithms used by the MCMC, and even to oneself write customized sampling algorithms for use in the MCMC. Indeed, the MCMC sampling algorithms used for fitting our models—specifically those used for the CRP concentration parameter $\alpha$, and for the CRP-distributed group membership indicators $g_{1:N}$—are themselves custom sampling algorithms written for precisely these non-parametric motifs, and added into nimble’s MCMC repertoire of algorithms. That said, nimble does not attempt to provide “canned” algorithms, nor any particular pre-written model structures, but rather an environment for writing custom functions, statistical algorithms, and distributions, and the application of these to generally-specified hierarchical model structures. The goal of nimble is to provide a flexible model and algorithmic programming environment to facilitate highly efficient analysis of models and complex data.

It is common that heterogeneity will be present to some degree in the detection process of ecological models. In practice, this may be detected by goodness-of-fit tests (Jeyam et al. 2020), or perhaps based on prior expert knowledge. When detection heterogeneity is known or suspected to be present, and suitable covariates are not available to accurately model this heterogeneity, we recommend using a BNP modeling approach. This approach alleviates the necessity of selecting the number of components used in a finite mixture model, which is an inherently difficult and oftentimes subjective process. No less, the exact number of mixture components is generally not the primary inferential focus. Use of a BNP modeling approach, as demonstrated herein, accounts for whatever degree or structure of detection heterogeneity that may be present. This provides an effective approach to reducing bias in the resulting ecological inferences.

**Author Contributions** All authors contributed to the methodology, analysis, and manuscript preparation.

**Data availability** All data used is available at github.com/danielturek/bnp-ecology-examples.

**Compliance with ethical standards**

**Funding** Support for DT was provided by Fulbright Research Scholarship Award 9183-FR and also the Williams College Class of 1945 World Fellowship. Support for CW was partially provided by award NSF-
DMS 1622444. Support for OG was provided by a grant from the French Centre National de la Recherche Scientifique (CNRS) and “Mission pour l’interdisciplinarité” through its “Osez l’interdisciplinarité call.” We warmly thank the French Office of Biodiversity (OFB) for sharing the wolf datasets.

Conflict of interest The authors have no conflicts of interest.

Code availability Code used is available at github.com/danielturek/bnp-ecology-examples.

References

Ahmed A, Xing E (2008) Dynamic non-parametric mixture models and the recurrent Chinese restaurant process: with applications to evolutionary clustering. In: Proceedings of the 2008 SIAM international conference on data mining. SIAM, pp 219–230

Begon M, Harper J, Townsend C (2006) Ecology: from individuals to ecosystems, 4th edn. Wiley-Blackwell, New York

Blackwell D, MacQueen J (1973) Ferguson distributions via Pólya urn schemes. Ann Stat 1:353–355

Blei DM, Griffiths TL, Jordan MI (2010) The nested Chinese restaurant process and Bayesian nonparametric inference of topic hierarchies. J ACM 57(2):1–30

Brooks S, Gelman A, Jones G, Meng X-L (2011) Handbook of Markov Chain Monte Carlo. CRC Press, New York

Cubaynes S, Lavergne C, Marboutin E, Gimenez O (2012) Assessing individual heterogeneity using model selection criteria: how many mixture components in capture–recapture models? Methods Ecol Evol 3(3):564–573

Cubaynes S, Pradel R, Choquet R, Duchamp C, Gaillard J-M, Lebreton J-D, Marboutin E, Miquel C, Reboulet A-M, Poillot C, Taberlet P, Gimenez O (2010) Importance of accounting for detection heterogeneity when estimating abundance: the case of French wolves. Conserv Biol 24(2):621–626

de Valpine P, Turek D, Paciorek CJ, Anderson-Bergman C, Lang DT, Bodik R (2017) Programming with models: writing statistical algorithms for general model structures with NIMBLE. J Comput Graph Stat 26(2):403–413

Diana A, Matechou E, Griffin J, Johnston A et al (2020) A hierarchical dependent Dirichlet process prior for modelling bird migration patterns in the UK. Ann Appl Stat 14(1):473–493

Dorazio RM, Mukherjee B, Zhang L, Ghosh M, Jelks HL, Jordan F (2008) Modeling unobserved sources of heterogeneity in animal abundance using a Dirichlet process prior. Biometrics 64(2):635–644

Escobar MD (1994) Estimating normal means with a Dirichlet process prior. J Am Stat Assoc 89:268–277

Ford J, Patterson T, Bravington M (2015) Modelling latent individual heterogeneity in mark-recapture data with dirichlet process priors. arXiv:1511.07103

Gelman A, Hwang J, Vehtari A (2014) Understanding predictive information criteria for Bayesian models. Stat Comput 24(6):997–1016

Gimenez O, Viallefont A, Charmantier A, Pradel R, Cam E, Brown CR, Anderson MD, Brown MB, Covas R, Gaillard J (2008) The risk of flawed inference in evolutionary studies when detectability is less than one. Am Nat 172(3):441–448

Goldstein BR, Turek D, Ponisio L, de Valpine P (2020) nimbleEcology: Distributions for ecological models in nimble. https://cran.r-project.org/package=nimbleEcology

Guillera-Arroita G (2017) Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. Ecography 40(2):281–295

Jeyam A, McCrea R, Pradel R (2020) Assessing heterogeneity in transition propensity in multistate capture–recapture data. J R Stat Soc Ser C 69(2):413–427

Li Y, Schofield E, Gönen M (2019) A tutorial on Dirichlet process mixture modeling. J Math Psychol 91:128–144

Link WA (2003) Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. Biometrics 59(4):1123–1130
Link WA (2004) Individual heterogeneity and identifiability in capture–recapture models. Anim Biodivers Conserv 27(1):87–91
Lo AY (1984) On a class of Bayesian nonparametric estimates I: density estimates. Ann Stat 12:351–357
Louvier J, Chambert T, Marboutin E, Gimenez O (2018) Accounting for misidentification and heterogeneity in occupancy studies using hidden Markov models. Ecol Model 387:61–69
Lunn D, Spiegelhalter D, Thomas A, Best N (2009) The BUGS project: evolution, critique and future directions. Stat Med 28(25):3049–3067
MacKenzie DI, Nichols JD, Royle JA, Pollock KH, Bailey L, Hines JE (2018) Occupancy estimation and modeling: inferring patterns and dynamics of species occurrence, 2nd edn. Elsevier, Amsterdam
Manrique-Vallier D (2018) Accounting for misidentification and heterogeneity in occupancy studies using hidden Markov models. Biometrics 74(2):1246–1254
Matechou E, Caron F (2017) Modelling individual migration patterns using a Bayesian nonparametric approach for capture–recapture data. Ann Appl Stat 11(1):21–40
McCrea RS, Morgan B (2014) Analysis of capture–recapture data, 1st edn. Chapman & Hall, New York
Neal RM (2000) Markov chain sampling methods for Dirichlet process mixture models. J Comput Graph Stat 9:249–265
NIMBLE Development Team (2019). Nimble: MCMC, particle filtering, and programmable hierarchical modeling. https://cran.r-project.org/package=nimble
Ohlssen DI, Sharples LD, Spiegelhalter DJ (2007) Flexible random-effects models using Bayesian semi-parametric models: applications to institutional comparisons. Stat Med 26(9):2088–2112
Pitman J (1995) Exchangeable and partially exchangeable random partitions. Probab Theory Relat Fields 102:145–158
Pitman J (1996) Some developments of the blackwell-Macqueen urn scheme. In: Statistics, probability, and game theory: papers in honor of David Blackwell. JSTOR, pp 245–267
Pledger S (2005) The performance of mixture models in heterogeneous closed population capture–recapture. Biometrics 61(3):868–873
Pledger S, Pollock KH, Norris JL (2010) Open capture-recapture models with heterogeneity: II. Jolly–Seber model. Biometrics 66(3):883–890
Pohle J, Langrock R, van Beest FM, Schmidt NM (2017) Selecting the number of states in hidden Markov models: pragmatic solutions illustrated using animal movement. J Agric Biol Environ Stat 22(3):270–293
Qin ZS (2006) Clustering microarray gene expression data using weighted Chinese restaurant process. Bioinformatics 22(16):1988–1997
Royle JA (2006) Site occupancy models with heterogeneous detection probabilities. Biometrics 62(1):97–102
Sethuraman J (1994) A constructive definition of Dirichlet prior. Stat Sin 2:639–650
Turek D, de Valpine P, Paciorek CJ (2016) Efficient Markov chain Monte Carlo sampling for hierarchical hidden Markov models. Environ Ecol Stat 23(4):549–564
Turek D, de Valpine P, Paciorek CJ, Anderson-Bergman C (2017) Automated parameter blocking for efficient Markov chain Monte Carlo sampling. Bayesian Anal 12(2):465–490
Warren CC, Veech JA, Wickerley FW, O’Donnell L, Ott JR (2013) Detection heterogeneity and abundance estimation in populations of golden-cheeked warblers (setophaga chrysopea). Auk 130(4):677–688
Watanabe S (2010) Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. J Mach Learn Res 11(12):3571–3594
Wehrhahn C, Rodriguez A, Paciorek C (2018) Bayesian nonparametric mixture models using nimble. In: NeurIPS workshop on nonparametric Bayesian models

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Daniel Turek is Assistant Professor of Statistics at Williams College. His research interests involve Bayesian sampling methodology and computational efficiency, with an applied interest in areas of ecology. He is also a core developer of the NIMBLE statistical software package.

Claudia Wehrhahn is Visiting Assistant Professor at University of California, Santa Cruz. Her research focuses on the development of Bayesian nonparametric models with applications in disease clustering and...
human mobility. She is also the main developer of the nonparametric functionality of the NIMBLE statistical software.

Olivier Gimenez is a senior researcher at CNRS in France. He is an ecological statistician working at the interface of animal ecology, statistical modeling and social sciences. He addresses questions in ecology and conservation biology dealing with the impact of climate change and human activities on population dynamics and species distribution and the management of large mammals.