BIOMETRIC METHODOLOGY

Biometrics WILEY
A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY

A general modeling framework for open wildlife populations based on the Polya tree prior

Alex Diana[1] | **Eleni Matechou[1]** | **Jim Griffin[2]** | **Todd Arnold[3]** | **Simone Tenan[4]** | **Stefano Volponi[5]**

[1]School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, UK

[2]Department of Statistical Science, University College London, London, UK

[3]Department of Fisheries, Wildlife and Conservation Biology, University of Minnesota, Minneapolis, Minnesota, USA

[4]Institute of Marine Sciences (CNR-ISMAR), Venezia, Italy

[5]Istituto Superiore Protezione e Ricerca Ambientale, Bologna, Italy

Correspondence
Alex Diana, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, UK.
Email: ad625@kent.ac.uk

Abstract

Wildlife monitoring for open populations can be performed using a number of different survey methods. Each survey method gives rise to a type of data and, in the last five decades, a large number of associated statistical models have been developed for analyzing these data. Although these models have been parameterized and fitted using different approaches, they have all been designed to either model the pattern with which individuals enter and/or exit the population, or to estimate the population size by accounting for the corresponding observation process, or both. However, existing approaches rely on a predefined model structure and complexity, either by assuming that parameters linked to the entry and exit pattern (EEP) are specific to sampling occasions, or by employing parametric curves to describe the EEP. Instead, we propose a novel Bayesian nonparametric framework for modeling EEPs based on the Polya tree (PT) prior for densities. Our Bayesian nonparametric approach avoids overfitting when inferring EEPs, while simultaneously allowing more flexibility than is possible using parametric curves. Finally, we introduce the replicate PT prior for defining classes of models for these data allowing us to impose constraints on the EEPs, when required. We demonstrate our new approach using capture–recapture, count, and ring-recovery data for two different case studies.

KEYWORDS
Bayesian nonparametrics, capture–recapture, count data, Polya tree, ring recovery, statistical ecology

1 | INTRODUCTION

In recent years, there has been an increased interest in monitoring wildlife populations due to the ongoing effects of climate change and habitat destruction. However, such monitoring is challenging because we cannot observe all animals present in the wild. Therefore, statistical models need to be employed to infer population sizes (Royle, 2004), migration (Pledger et al., 2009), phenological (Dennis et al., 2017), survival (McCrea et al., 2013), or entry and exit patterns (EEPs) (Matechou et al., 2016) from ecological data at one or more sites.

Although these models are developed for data collected using different sampling protocols, they all focus on the estimation of the EEPs of populations, where entry can correspond to arrival/birth, exit to departure/death, and corresponding length of stay (LOS) to retention/survival at the site. These existing methods either assume that

parameters related to EEPs are completely unconstrained, requiring two parameters to be estimated for each sampling occasion, one for entry and one for exit (Lyons et al., 2016; Pledger et al., 2009), or use parametric curves to constrain entry (Matechou et al., 2014) and exit (Jimenez-Muñoz et al., 2019; McCrea et al., 2013) patterns. However, in the former case, the model can be overparameterized and difficult to interpret, whereas in the latter case, the degree of smoothness is predefined and hence the model may not be flexible enough.

In recent years, Bayesian nonparametric (BNP) ecological models have been developed as an option for flexibly modeling EEPs without making parametric assumptions. An attractive feature of BNP models is that they can be defined such that the prior predictive distribution (or, equivalently, the prior mean) is given by a parametric distribution, referred to as the centering distribution. By centering a BNP model over a parametric distribution, we obtain a posterior distribution that is a compromise between a fully unconstrained model and the parametric model used to specify the centering distribution. In particular, models have been built using the Dirichlet Process mixture prior (Lo, 1984) (see, e.g., Diana et al., 2020; Ford et al., 2015; Matechou & Caron, 2017), and the Polya tree (PT) prior (Ferguson, 1974; Lavine, 1994) (see, e.g., Diana et al., 2018). The Dirichlet process mixture defines a prior for a distribution as an infinite mixture, whereas the PT prior is constructed by recursively assigning probability mass to sequences of nested partitions of the sample space, which follow a tree structure.

In ecological surveys, data are typically collected at discrete observation times, referred to as sampling occasions, and every pair of consecutive sampling occasions defines an interval, during which individuals may enter or exit the study area. No data are available between sampling occasions, which suggests that EEPs should be modeled on the grid defined by the sampling occasions. In this paper, we develop a general framework for modeling EEPs based on the PT prior using this grid, extending the approach of Diana et al. (2018). The use of the PT allows us to build a model directly on the distributions of EEPs, with minimal parametric assumptions on the shape of these distributions. Additionally, as we demonstrate in this paper, the PT prior can lead to positive correlation between consecutive entry (exit) intervals (Lavine, 1994). This results in a smoother EEP than unconstrained models with the degree of smoothness informed by the data instead of being predefined by parametric curves. Our proposed PT approach leads to more efficient inference methods compared to existing BNP approaches based on Dirichlet process mixtures, which introduce individual entry and exit times as continuous latent variables in Markov chain Monte Carlo (MCMC) samplers. Instead, with our PT approach, only the number of individuals in each cell of the grid needs to

be inferred, which means that the number of latent variables grows with the number of sampling occasions and not with the number of individuals in the population.

Finally, we introduce the replicate PT (RPT) framework, which allows us to impose constraints on the parameters of the entry (exit) process as required, by replicating parts of the tree. These constraints are similar to those typically employed in ecological models, such as constraining the distribution of LOS to be the same for all entry times, leading to more parsimonious models with interesting ecological interpretation. For example, we use the RPT to assume age-dependence when modeling the LOS distribution in the context of ring-recovery (RR) data and to assume time-dependence when modeling the exit distribution in the context of capture–recapture (CR) data.

We demonstrate our novel framework using two different case studies. The first considers the estimation of EEPs of Eurasian Spoonbills (*Platalea leucorodia*) using a joint model of CR and count data collected on the same population. The second considers the estimation of age-specific survival rates using RR data of a population of mallards (*Anas platyrhynchos*), where only some of the birds are of known age, and demonstrates the advantage of the PT in modeling survival from sparse data, without resorting to the use of parametric curves.

The paper is structured in the following way. In Section 2, we describe the main features of the PT prior and use a time-to-event example as an illustration of univariate partitions. In Section 3, we introduce the RPT and use it to define constrained versions of the time-to-event example. In Section 4, we define bivariate partitions, whereas in Section 5, we summarize the Diana et al. (2018) model, we define the joint model for CR data and count data and the model for RR data. The models are demonstrated using two case studies in Section 6, whereas Section 7 concludes the paper and introduces some potential future directions.

2 | THE POLYA TREE PRIOR

The PT prior is a nonparametric prior for a probability distribution G with sample space Ω. The PT has two parameter sets: the first is a sequence of nested partitions Π of the sample space Ω, whereas the second parameter, α, is a sequence of positive numbers associated with each set of each partition. The partition at the first level, π_1, is obtained by splitting the sample space into two sets, $\{B_0, B_1\}$. Subsequently, to build the partition at the second level, π_2, each set B_i is split into two sets, $\{B_{i0}, B_{i1}\}$, next each set B_{ij} is split into the sets $\{B_{ij0}, B_{ij1}\}$, and so on.

The PT recursively assigns a prior to $G(B)$ for all B's in the sequence of partitions according to the following procedure. By defining $\epsilon_1 \ldots \epsilon_m$ as a generic sequence of 0s and 1s, with $m \in \mathbb{N}^+$, and $B_{\epsilon_1 \ldots \epsilon_m}$ as the corresponding set of the

partition, the random mass associated with $B_{\epsilon_1...\epsilon_m}$ is given by

$$G(B_{\epsilon_1...\epsilon_m}) = V_{\epsilon_1...\epsilon_m} G(B_{\epsilon_1...\epsilon_{m-1}}) = \prod_{i=1}^{m} V_{\epsilon_1...\epsilon_i}, \quad (1)$$

where $V_{\epsilon_1...\epsilon_i 0} \sim \text{Beta}(\alpha_{\epsilon_1...\epsilon_i 0}, \alpha_{\epsilon_1...\epsilon_i 1})$ and $V_{\epsilon_1...\epsilon_i 1} = 1 - V_{\epsilon_1...\epsilon_i 0}$. It follows that, if $y \sim G$ where $G \sim \text{PT}$, then $\mathbb{P}(y \in B_{\epsilon_1...\epsilon_m} | y \in B_{\epsilon_1...\epsilon_{m-1}}) = \frac{G(B_{\epsilon_1...\epsilon_m})}{G(B_{\epsilon_1...\epsilon_{m-1}})} = V_{\epsilon_1...\epsilon_m}$, so that each $V_{\epsilon_1...\epsilon_m}$ is the conditional probability of y falling in set $B_{\epsilon_1...\epsilon_m}$ given that y falls in the set $B_{\epsilon_1...\epsilon_{m-1}}$ or the proportion of $G(B_{\epsilon_1...\epsilon_{m-1}})$ allocated to $G(B_{\epsilon_1...\epsilon_m})$. We note that in the case of nondyadic splits, that is, in the case where each set is split into more than two sets, the PT can be defined analogously by replacing the Beta distributions with Dirichlet distributions. In its standard form, the PT is defined for an infinite sequence of partitions. However, in this paper, we assume a PT prior for the distributions of EPPs, which cannot be inferred at a resolution finer than the times of observation, that is, at the times when the K sampling occasions take place, $t_1, ..., t_K$, hence in what follows we only define finite partitions.

The partition can be chosen according to the specific application. For example, suppose that we are interested in the distribution of continuous time-to-event data but only observe events at times $t_1, ..., t_K \in \mathbb{R}$, that is, we only know in which of the intervals $(0, t_1], ..., (t_K, \infty)$ the event occurred. In this case, using an idea similar to Lavine (1994), we can build a partition by splitting the sample space $\Omega = (0, \infty)$ into $B_0 = (0, t_1]$ and $B_1 = (t_1, \infty)$ at the first level, with B_1 further split into $B_{10} = (t_1, t_2]$ and $B_{11} = (t_2, \infty)$ at the second level, and so on, until at the last level $B_{\mathbf{1}_{K-1}} = (t_{K-1}, \infty)$ is split into $B_{\mathbf{1}_{K-1}0} = (t_{K-1}, t_K]$ and $B_{\mathbf{1}_{K-1}1} = (t_K, \infty)$ where $\mathbf{1}_j = \underbrace{(1, ..., 1)}_{j \text{ times}}$. In this way, the PT assigns a prior for the masses of the sets $(0, t_1], (t_1, t_2], ..., (t_K, \infty)$. The structure of the tree allows us to define a simpler notation for the V's so that $V_j = \mathbb{P}(x \in B_{\mathbf{1}_j 0} | x \in B_{\mathbf{1}_j})$ (and by convention, $V_0 = \mathbb{P}(x \in (0, t_1])$). Hence, for example, $\mathbb{P}(x \in B_{110}) = \mathbb{P}(x \in B_1)\mathbb{P}(x \in B_{11} | x \in B_1)\mathbb{P}(x \in B_{110} | x \in B_{11}) = (1 - V_0)(1 - V_1)V_2$. A representation of the scheme is shown in Figure 1 and we refer to this partition scheme as the *forward partition*. Examples of samples from a forward partition are shown in Figure S1 of the Supporting Information.

The PT can be centered on any distribution G_0 in the sense that, for every set B of the partition, $\mathbb{E}[G(B)] = G_0(B)$. This is achieved by setting the αs in definition (1) such that

$$\frac{\alpha_{\epsilon_1...\epsilon_i 0}}{\alpha_{\epsilon_1...\epsilon_i 0} + \alpha_{\epsilon_1...\epsilon_i 1}} = \mathbb{E}[V_{\epsilon_1...\epsilon_i 0}] = \frac{G_0(B_{\epsilon_1...\epsilon_i 0})}{G_0(B_{\epsilon_1...\epsilon_i})},$$

for all $\epsilon_1 ... \epsilon_i$. We note that this defines the αs only up to proportionality, as the proportionality constant is defined by the variance around the centering distribution. In the case above of time-to-event data and a forward partition, the α_s are defined such that $\frac{\alpha_0}{\alpha_0 + \alpha_1} = \mathbb{E}[V_0] = \frac{G_0((0, t_1])}{G_0([0, \infty))}$, $\frac{\alpha_{10}}{\alpha_{10} + \alpha_{11}} = \mathbb{E}[V_1] = \frac{G_0((t_1, t_2])}{G_0((t_1, \infty))}$, and so on. For example, if a Weibull(λ, k) distribution is taken as the centering distribution, as in the case of the RR case study presented in Section 5.3, then $G_0((a, b)) = e^{-(\frac{a}{\lambda})^k} - e^{-(\frac{b}{\lambda})^k}$, and $\frac{\alpha_0}{\alpha_0 + \alpha_1} = 1 - e^{-(\frac{t_1}{\lambda})^k}$, $\frac{\alpha_{10}}{\alpha_{10} + \alpha_{11}} = e^{-(\frac{t_1}{\lambda})^k} - e^{-(\frac{t_2}{\lambda})^k}$, and so on.

The following proposition, proven in the Supporting Information, is useful for understanding how the choice of partition affects the smoothness of a sample from a PT.

Proposition 1. *Let G be a sample from a PT using the forward partition, where $V_i \sim \text{Beta}(\alpha_i, \alpha_i)$ and let $Y_i = G([t_i, t_{i+1}))$. If α_i is an increasing sequence of positive numbers, it follows that $Cor(Y_i, Y_{i+1}) < Cor(Y_{i+1}, Y_{i+2})$.*

Proposition 1 shows that if the distributions of the random variables (RVs) $V_0, V_1, ...$ have parameters whose sum is not decreasing, the correlation between the masses of adjacent intervals increases for intervals corresponding to later times or, equivalently, that the distribution of the V's becomes smoother toward the right tail. This is the case for the forward partition, defined above. Conversely, we can assume that the distribution of the Vs is smoother in the left tail, by reversing the order of the split so that $B_0 = [t_K, \infty)$ and $B_1 = (0, t_K)$ at the first level, $B_{10} = [t_{K-1}, t_K)$ and $B_{11} = (0, t_{K-1})$ at the second level, and so on. We term this partition scheme the *backward partition*.

For ease of reference, we list below the univariate partitions on which we are going to rely for the rest of the paper. In all cases, the partition is defined using the observation times $t_1, ..., t_K$. We note that by convention, we define $t_0 = 0$ and $t_{K+1} = \infty$.

- *Forward partition:* this partition corresponds to the idea of Lavine (1994), with the first split defined using t_1 and the last using t_K.
- *Backward partition:* this partition is analogous to the forward partition but now the first split is defined using t_K and the last using t_1.
- *Uniform partition:* this partition is built in a single step by splitting using all the observation times simultaneously. That is, Ω is split into $(B_0, ..., B_K)$, where $B_i = (t_i, t_{i+1}]$, resulting in a tree that has a single level. Since this split is not dyadic, the masses are assigned according to a K-dimensional Dirichlet distribution with parameters $\alpha G_0(B_0), ..., \alpha G_0(B_K)$. We note that, in contrast to
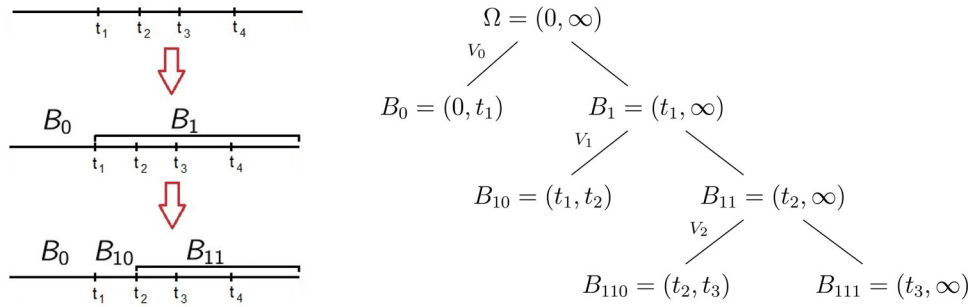
FIGURE 1 Sequence of splits (left) and structure (right) of the partition for the time-to-event data example. The random variable assigning the ratio of the masses of two sets for each branch is represented on the left branch. The mass assigned to the set in each right branch is always one minus the mass assigned to the set in the corresponding left branch.

the forward and backward partitions, the uniform partition imposes a negative correlation between the masses of any two intervals, which can be useful in situations where entry and/or exit patterns are not expected to be smooth.

2.1 | Example: Time-to-event

In this section, we demonstrate how to apply the PT prior to the estimation of survival probabilities using time-to-event data. We assume that each individual is followed starting at a corresponding time in $\{t_1, \ldots, t_K\}$, referred to as time of first observation (TFO), when they are of age 1, until their death, which is observed at a subsequent time.

A common modeling approach is to assume that individuals of age a at time t_j remain alive until time t_{j+1} with probability $\phi_{j,a}$. To formulate this approach in a PT framework, we first arrange all the followed individuals in a set of vectors $\mathbf{n}^k = (n_1^k, \ldots, n_{K-k+1}^k)$, where n_j^k indicates the number of individuals with TFO equal to t_k and with death in the interval (t_{k+j-1}, t_{k+j}), that is, observed at time t_{k+j}. Next, we assume a distribution G_k on the time of death of the individuals belonging to vector $\mathbf{n}^k$. The sample space for G_k is $\Omega_k = (t_k, \infty)$, as time of death is left truncated by the TFO for each individual. We assume independent PT priors for the G_k's, with partition taken to be the forward partition, according to which the sample space Ω_k is split into $B_0^k = (t_k, t_{k+1}]$ and $B_1^k = (t_{k+1}, \infty)$, next B_1^k is further split into $B_{10}^k = (t_{k+1}, t_{k+2}]$ and $B_{11}^k = (t_{k+2}, \infty)$, and so on. The last level of the partition corresponds to the sets $(t_k, t_{k+1}], \ldots, (t_K, \infty)$, and hence, the PT prior for G_k defines a prior for the probability that death occurs in each of these intervals. We note that the K partitions have different depths, as for individuals with TFO equal to t_k, the partition is built only for $K - k$ levels.

In this example, each G_k is defined by $V_0^k, \ldots, V_{K-k-1}^k$, where $V_j^k = \frac{G_k(B_{1_j0})}{G_k(B_{1_j})}$. Therefore, V_j^k corresponds to the

probability that an individual with TFO t_k is not present at time t_{k+j+1} conditional on being present at time t_{k+j}, that is, $V_j^k = 1 - \phi_{k+j,j+1}$. Finally, if m_k is the number of individuals with TFO t_k, the model for vector $\mathbf{n}^k$ is given by

$$n_1^k \sim \text{Bin}(m_k, V_0^k),$$

$$n_j^k \sim \text{Bin}\left(m_k - \sum_{l=1}^{j-1} n_l^k, V_{j-1}^k\right), \qquad j = 2, \ldots, K-k. \quad (2)$$

3 | REPLICATE PT

In the example discussed in Section 2.1, the parameters V_j^k belong to independent PTs and hence the probability of survival varies freely by time and age. To build simpler models, we need to reduce the number of parameters by introducing constraints on the V_j^k's within our RPT framework. For example, the number of parameters can be reduced by assuming that the probability of survival is either age-dependent or time-dependent, which is standard practice in the ecology literature, leading to a more parsimonious model.

In the PT framework, assumptions of this kind can easily be incorporated into the model by sharing RVs across different PTs or, in other words, via "replicating" tree structures across different trees. This idea will also be useful later when defining models for CR and RR data, and hence, we formalize it with the following definition. For simplicity, we provide the definition for two trees but it is straightforward to extend it to multiple trees.

Let Σ be a rule to generate a sequence of partitions from an initial seed set. For example, the rule Σ can correspond to the forward partition. Let G_1 and G_2 be distributions having PT priors with partition Π_1 and Π_2, respectively, with V^1 and V^2 the sets of RVs corresponding to G_1 and G_2, respectively, where $V_{\epsilon,0}^l = \frac{G_l(B_{\epsilon 0})}{G_l(B_\epsilon)}$ as defined in Section 2. We say that G_1 and G_2 have an RPT structure across

two sets $B^1_\epsilon \in \Pi_1$ and $B^2_{\epsilon'} \in \Pi_2$ if the following constraints hold:

- The partitions of the trees starting from B^1_ϵ and $B^2_{\epsilon'}$ are generated according to the same rule Σ, although the two partitions could be stopped after a different number of steps;
- For all $\epsilon_1, \dots, \epsilon_m$, $V^1_{\epsilon\epsilon_1\dots\epsilon_m} = V^2_{\epsilon'\epsilon_1\dots\epsilon_m}$.

The first condition states that the partitions of the trees starting from B^1_ϵ and $B^2_{\epsilon'}$ are the same (even if they might not have the same depth), whereas the second states that the RVs in the two trees are the same. We note that the definition can also be used with $B^1_\epsilon \in \Pi_1$ and $B^1_{\epsilon'} \in \Pi_1$ that allows sharing across the same tree.

We can use the RPT to impose different constraints on the probability of survival for the time-to-event example of Section 2.1, which we list below.

- *Unconstrained case*: The most general case is obtained by assuming that all the V^k_j are different. In this case, the probability of survival varies by age and by time and no replicate structure is assumed (Figure S4 of the Supporting Information).
- *Age-dependent case*: $\phi_{\cdot,a} \equiv \phi_a \Rightarrow V^k_j \equiv V_j \ \forall k$. This is equivalent to assuming that $G_k \equiv G \ \forall k$. In the RPT framework, this is equivalent to requiring that the G_k have a replicate structure across the sets $\Omega_1, \dots, \Omega_K$ (Figure S2 of the Supporting Information).
- *Time-dependent case*: $\phi_{j,\cdot} \equiv \phi_j \Rightarrow V^k_j \equiv V_{k+j}$. This is equivalent to assuming that the distribution of $G_k(x)$ is the same as the distribution of $G_{k-1}(x | x \in B_{k-1,1}) \ \forall k$. In the RPT framework, this is obtained assuming a replicate structure across the pairs of trees with seed sets: Ω_2 and B^1_1, Ω_3 and B^2_1, Ω_4 and B^3_1, and so on (Figure S3 of the Supporting Information).
- *Constant case*: $\phi_{j,a} \equiv \phi \Rightarrow V^k_j \equiv V \ \forall k, j$. This is equivalent to assuming that the probability of survival is always constant. In this case, all the trees collapse to a single RV.

We note that employing constraints of the type induced by the RPT structure gives rise to models that assume independence between exit/LOS and TFO. This idea is extended further in Section 5. For ease of exposition, in all the case studies shown later, we only consider the most commonly employed constraint in each case. However, different RPT structures can be employed to assume different constraints as required.

4 | BIVARIATE PARTITIONS

To define joint models for entry and exit, or entry and LOS, we need to introduce partitions for distributions on

$\mathbb{R}^2$. These partitions can be built sequentially using the schemes described in Section 2.

A useful bivariate partition can be constructed by first applying the forward partition in one dimension (entry) and next the backward partition in the other dimension (exit), as shown in Figure 2. We call this the *entry and exit partition*, and we give a formal definition in Section 3.1 of the Supporting Information. This partition is useful when we expect the entry pattern to be less smooth in the left tail compared to the right, with the opposite being true for the exit pattern. Examples of this EEP are stopover sites (Matechou et al., 2013) or breeding sites (Diana et al., 2020), where most individuals enter at the start of the study period and exit at the end, with entry and exit taking place in waves, and hence the entry (exit) pattern being more spiky at the start (end) and smoother at the end (start).

The entry and exit partition allows us to elicit prior information on the entry and exit distributions. However, in some cases, prior information is available on the LOS distribution instead. For eliciting a prior on the LOS distribution, we define a bivariate partition based on the forward partition, by first splitting the sample space with respect to the LOS and then with respect to the entry intervals. We describe how to construct this partition in the following paragraph.

First, we split the sample space into the set $B_0 = \{(x, y) : t_k < x < y < t_{k+1} \ \forall k\}$, consisting of the individuals exiting immediately after entry and the set $\overline{B}_0 = \{(x, y) : t_k < x < t_{k+1} \leq y \ \forall k\}$, consisting of the remainder of the individuals (Level 1 of Figure 3). Next, we split $\overline{B}_0$ into the set $B_1 = \{(x, y) : t_k < x < t_{k+1} < y < t_{k+2} \ \forall k\}$ of individuals staying for only one interval and the set $\overline{B}_1 = \{(x, y) : t_k < x < t_{k+1}, y \geq t_{k+2} \ \forall k\}$ of individuals remaining for more than one interval (Level 2 of Figure 3). The process is repeated K times for the remaining observation times. Next, each set is split according to the entry intervals using the uniform partition. This partition is termed *entry and LOS partition* and an application is presented in Section 5.3.

5 | ECOLOGICAL MODELS

In this section, we define ecological models for count, CR and RR data. These data are collected according to different survey methods, which are all prone to false-negative observation errors, and hence, not every individual that is present on a sampling occasion will be observed on that occasion. Therefore, each individual has a latent *presence history*, which is a vector with entries equal to 1 on sampling occasions when the individual was present and 0 otherwise. The observation process for each survey type is defined conditional on these latent presence histories.
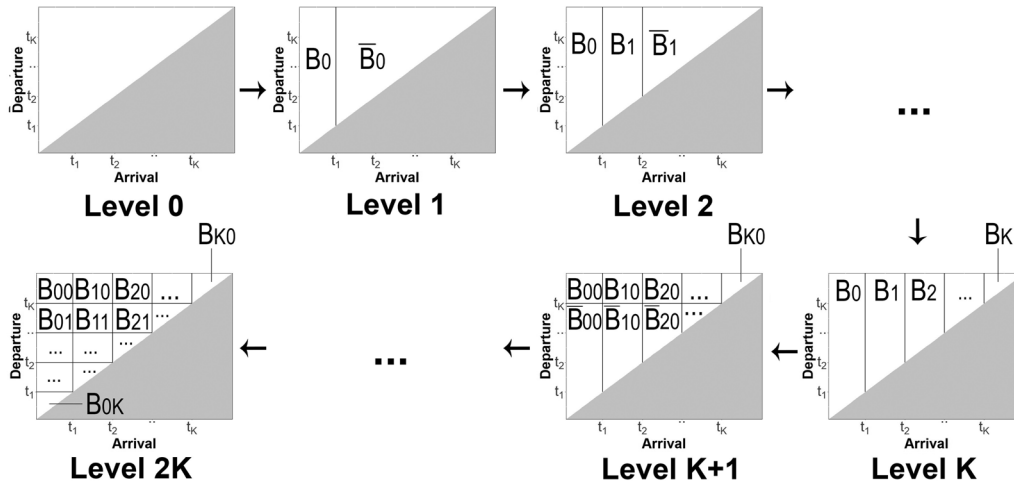
FIGURE 2 Entry and exit partition. Partition of a PT prior for a distribution defined on $\{(x, y) : (x, y) \in \mathbb{R}^2, y > x\}$, built by first using the forward partition and next using the backward partition. The first K steps are performed using the forward partition in the entry dimension, and the next K steps using the backward partition in the exit dimension. The shaded area represents the region excluded from the sample space, because departure is left truncated by arrival. The ticks on the x and y-axes represent the sampling occasions.
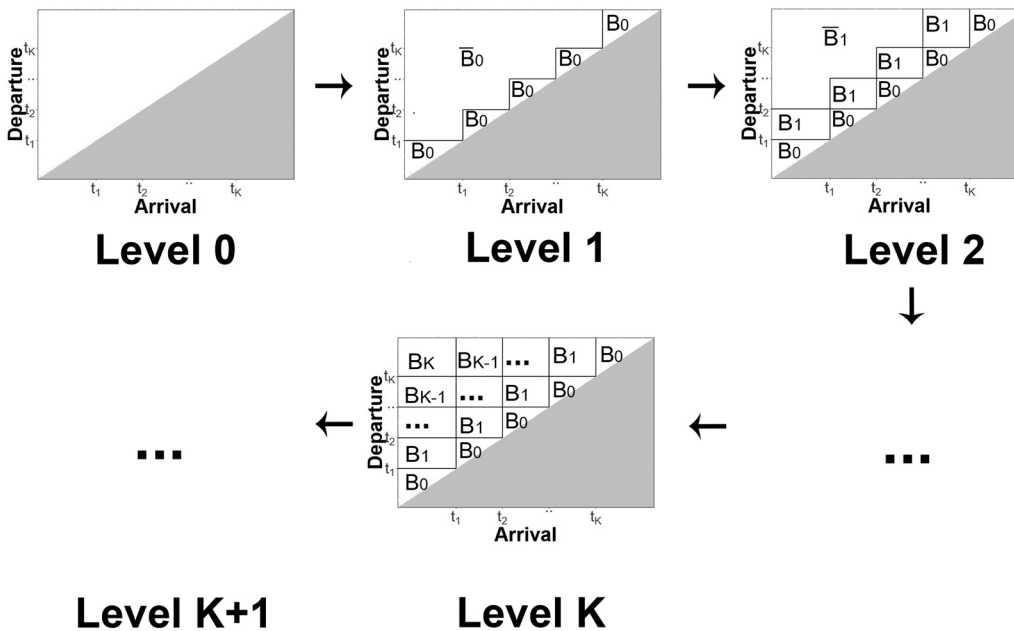
FIGURE 3 Entry and LOS partition. Partition of a PT prior for a distribution defined on $\{(x, y) : (x, y) \in \mathbb{R}^2, y > x\}$ built using the entry and LOS partition. The first K levels are built splitting according to the LOS, whereas the $K + 1$th split is performed according to the entry dimension. The shaded area represents the region excluded from the sample space, because departure is left truncated by arrival. The ticks on the x and y-axes represent the sampling occasions.

Hence, in this section, we extend the ideas introduced in Section 2.1, where the time-to-event was observed for each individual, to modeling ecological data where the entry and exit times for each individual, as well as the size of the population, are unobservable. Therefore, within this framework, we infer the number of individuals in each cell of the grid defined in Section 4, and as a result also infer the population size.

5.1 | Example: Count data model

In this section, we briefly summarize the model of Diana et al. (2018) developed for count data as an illustration of the entry and exit partition for modeling ecological data. Count data are collected by visiting the site and simply detecting a number of individuals, without attempting to uniquely identify them. The data can be summarized in

a vector $(C_1, \dots, C_K)$, where C_k denotes the number of individuals detected on sampling occasion k.

The latent presence histories can be summarized in a matrix $\{n_{fl}\}_{f,l=0,\dots,K}$, where n_{fl} is the number of individuals with entry interval (t_f, t_{f+1}) and exit interval (t_l, t_{l+1}). The number of individuals available for detection on sampling occasion k, N_k, is readily available from the matrix n_{fl} since $N_k = \sum_{l=k}^{K} \sum_{f=0}^{k-1} n_{fl}$. Assuming that each individual can be detected with probability p independently of the other individuals, the observation process can be expressed as $C_k \sim \text{Bin}(N_k, p)$, $k = 1, \dots, K$.

A prior distribution is defined on the probability ω_{fl} that an individual belongs to cell n_{fl} using a PT prior distribution with the entry and exit bivariate partition of Section 4. We define X and Y to be the entry and exit time and let $V_i = \mathbb{P}(X \in (t_i, t_{i+1}]|X \in (t_i, \infty))$, for $i = 0, \dots, K-1$, and $W_j^i = \mathbb{P}(Y \in (t_{K-j}, t_{K-j+1}]|Y \in (t_i, t_{K-j+1}], X \in (t_i, t_{i+1}])$, for $j = 0, \dots, K-i-1$. The probabilities ω_{fl} can be obtained as a product of V's and W's as in definition (1). For example, ω_{23} is equal to $(1-V_0)V_1(1-W_0^2)W_1^2$, where the first two terms represent the probability of entering in $(t_2, t_3]$ and the last two terms represent the probability of exiting in $(t_3, t_4]$ conditional on entering in $(t_2, t_3]$.

Assuming a Poisson(ω) prior distribution on the population size N, with $N = \sum_{f=1}^{K} \sum_{l=1}^{K} n_{fl}$, the model can be summarized as

$$n_{fl}|\omega_{fl} \overset{ind.}{\sim} \text{Poisson}(\omega \times \omega_{fl}), \quad \{\omega_{fl}\}_{f,l=0,\dots,K} \sim \text{PT}(\Pi, \alpha),$$

where the expression on the right summarizes the prior distribution assumed for ω_{fl} through the V's and W's.

5.2 | Capture–recapture

In this section, we demonstrate how to adapt the model for count data described in the previous section to CR data using an RPT. In the case of CR data, on each sampling occasion, a number of individuals are caught, and newly caught individuals are uniquely marked. If D is the total number of captured individuals, the captures can be summarized in a $D \times K$ matrix, H, where H_{ik} is 1 if the ith individual was captured on sampling occasion k and 0 otherwise. Each row of the matrix is called a capture history.

To represent the latent presence histories in this case, we define the set of matrices $\{\mathbf{n}^k\} = \{n_{fl}^k\}_{f=1,\dots,K+1,l=1,\dots,K+1}$, where n_{fl}^k denotes the number of individuals with first possible capture at time t_f, first capture at time t_k and last possible capture at time t_l (clearly, $n_{fl}^k = 0$ for $f > k$). For $k = 0$, we define $\mathbf{n}^0$ as the matrix containing the individuals never captured. The likelihood of the individ-

uals with first capture on sampling occasion k can be written explicitly given $\mathbf{n}^k$ (Section 4 of the Supporting Information).

We denote by D_k the sum of the elements in each matrix $\mathbf{n}^k$, and we note that D_k is known for $k \neq 0$, as it is equal to the number of individuals with first capture at time t_k, whereas D_0 is unknown, as it is equal to the number of individuals never captured. Finally, the population size, N, is obtained as $N = \sum_{k=0}^{K} D_k$. For each matrix $\mathbf{n}^k$, we define a distribution G_k on the probabilities ω_{fl}^k of an individual belonging to each cell of the matrix $\{n_{fl}^k\}$. Assuming a Poisson(ω) prior distribution on N, the model can be written as

$$n_{fl}^k \sim \text{Multinomial}(D_k, \omega_{fl}^k), \quad D_0 + \sum_{k=1}^{K} D_k \sim \text{Poisson}(\omega).$$

We assume a PT prior for each G_k using the bivariate entry and exit partition of Section 4. We make two assumptions on the probabilities ω_{fl}^k, which will be employed using specific RPT structures. First, we assume that the entry and exit distribution is the same for individuals first captured on different sampling occasions. However, we note that each matrix is defined on a different sample space, as $\mathbf{n}^k$ is defined for $f \leq k, l > k$, as individuals in the kth matrix have to enter before the kth sampling occasion and exit after. A representation of one matrix is shown in Figure S5 of the Supporting Information. Therefore, we assume that the hazards of the distributions G_k are the same, that is, if X and Y are the entry and exit time of an individual first captured on sampling occasion k, the RVs $V_i^k = \mathbb{P}(X \in (t_i, t_{i+1}]|X \in (t_i, \infty))$, $i = 0, \dots, k-1$, and $(W_j^i)^{(k)} = \mathbb{P}(Y \in (t_{K-j}, t_{K-j+1}]|Y \in (t_i, t_{K-j+1}], X \in (t_i, t_{i+1}])$, $j = 0, \dots, K-k$, are the same for all k.

We also make the assumption that the exit distribution is time-dependent, in the sense that the exit interval of an individual does not depend on their entry interval, in a similar way to Section 3. This is achieved by imposing a RPT structure for each distribution G_k across the sets $B_0 = \{(x, y) : x \leq t_1\}$, $B_1 = \{(x, y) : t_1 < x \leq t_2\}$, ..., $B_{k-1} = \{(x, y) : t_{k-1} < x \leq t_k\}$ (see level K of Figure 2 for a visual representation of the sets). With reference to the notation of V and W introduced in Section 5.1, this is equivalent to assuming that $W_j^i \equiv W_j$. This assumption leads to a substantial reduction in the number of parameters but can be relaxed if required by not employing the RPT structure leading to different set of variables W_j^i, $j = 0, \dots, K-k$, for each time of entry $i = 0, \dots, K$.

We note that an alternative approach to write the likelihood explicitly is to model the entry and exit intervals of each individual, as opposed to using the grid

approach. This alternative approach is useful when the number of captured individuals is low compared to the number of sampling occasions, because it can lead to a reduced number of latent variables. This is the case for the case study considered in this paper, because there are only 40 marked individuals and eight sampling occasions. Section 5 of the Supporting Information presents the expression for the number of latent variables in each case.

5.3 | Ring recovery

According to the RR protocol, individuals are captured and marked each year and, after being marked, they can be recovered soon after they die. The data are summarized in an upper-triangular $K \times K$ matrix, R, where cell (i, j) denotes the number of individuals marked in year i and recovered dead in year j, and in a vector $(m_1, \dots, m_K)$ where m_k denotes the total number of individuals marked in year k. In the following, to be consistent with the notation used in the previous section, we will refer to the years as sampling occasions.

Similarly to the CR case, to jointly model EEPs, we work with a set of matrices $\mathbf{n}^k = \{n_{fl}^k\}$, where in this case, n_{fl}^k represents the latent number of individuals marked on sampling occasion k that were present in the population for f sampling occasions before being marked and remained for l sampling occasions after being marked. We assume that individuals could have entered the population for up to U sampling occasions before they were marked and could have stayed for up to U sampling occasions after being marked, which is equivalent to assuming that the dimension of each matrix $\mathbf{n}^k$ is $(U + 1) \times (U + 1)$. We have made this assumption of symmetry for simplicity. The choice of U is not critical as U is only used to fix the dimension of the grid and serves as an upper bound for the lifespan of the individuals. Therefore, as long as U is sufficiently larger than the lifespan of the individual, results are not sensitive to this choice.

We assume that an individual is recovered on the sampling occasion immediately following the interval in which they died with probability λ (and that the individual cannot be subsequently recovered). The number of individuals marked at time t_k that can be recovered at time t_{k+j} can be obtained by the matrix n_{fl}^k as $\sum_{f=0}^{U} n_{fj}^k$ by summing over the different entry times. Hence, the number of individuals marked at time t_k and recovered dead at time t_{k+j}, $R_{k,k+j}$, is distributed as Binomial($\sum_{f=0}^{U} n_{fj}^k, \lambda$).

We assume a PT prior for the probabilities ω_{fl}^k of an individual belonging to the cell n_{fl}^k, with partition taken to be the entry and LOS partition built in Section 4, and we

define as G^k the distribution of the probabilities of each cell for the matrix $\mathbf{n}^k$.

According to the entry and LOS partition, first, the sample space is split into B_0 and $\overline{B}_0$ (Level 1 of Figure 4), that is, the individuals staying for 0 sampling occasions and those staying for more than 0. Next, $\overline{B}_0$ is split into B_1 and $\overline{B}_1$ (Level 2 of Figure 4), that is, the individuals staying for 1 sampling occasion and staying for more than 1, and so on, for $2U$ levels. The first $2U$ steps split the sample space according to the LOS. Similarly to Section 2.1, we define as V_j^k the probability that an individual marked on sampling occasion k dies after $j + 1$ sampling occasions conditional on being alive for j sampling occasions after k. Therefore, $V_j^k = \frac{G^k(B_j)}{G^k(\overline{B}_{j-1})}$ (with $\overline{B}_{-1} = \Omega$).

At the latest level (level $2U$ of Figure 4), we use the uniform partition to split each diagonal according to the entry (or exit) intervals. To perform this step, we generate a set of RVs $(W_0^k, \dots, W_U^k) \sim \text{Dirichlet}(\alpha_0, \dots, \alpha_U)$ and we assign the masses of diagonal j, $(B_{j0}, \dots, B_{jj})$, proportionally to $W_0^k, \dots, W_U^k$, that is, as

$$G(B_{ji}) = G(B_j) \begin{cases} \frac{W_i^k}{W_0^k + \dots + W_j^k} & i = 0, \dots, j \qquad j = 0, \dots, U \\ \frac{W_{i+j-U}^k}{W_{j-U}^k + \dots + W_U^k} & i = 0, \dots, 2U - j \qquad j = U + 1, \dots, 2U. \end{cases}$$

The RVs W_j^k can be interpreted as the relative probabilities of entering in each interval previous to the time of marking.

Similarly to the CR case in Section 5.2, an assumption needs to be made about how the RVs W_j^k, $j = 1, \dots, K$ vary across the time of marking k. In the case of RR data, we cannot model the process by which individuals are first caught but instead only the recruitment of marked individuals. Hence, we choose to make the assumption that regardless of the time of marking, individuals of unknown age are equally likely to have entered l, $l = 0, \dots, U$, intervals before being marked. This is equivalent to assuming that the proportion of marked individuals of each age group is constant for each sampling occasion, in a similar way to McCrea et al. (2013). This assumption is imposed by assuming that $W_j^k \equiv W_j$, $k = 1, \dots, K$.

The second assumption we make is that the survival probability is age-dependent, that is, the probability of an individual surviving until the next sampling occasion depends only on the age of the individual and not on the sampling occasion the individual was marked. If we define Ω_k to be the sample space of the distribution G_k modeling the individuals marked at time t_k, this assumption can be achieved by assuming for $G_1, \dots, G_K$ an RPT across $\Omega_1, \dots, \Omega_K$. This forces the probabilities ω_{fl}^k to be the same for each k, that is, $\omega_{fl}^k \equiv \omega_{fl}$. This is equivalent to assuming that $V_j^k \equiv V_j$ or, similarly to the time-to-event model,
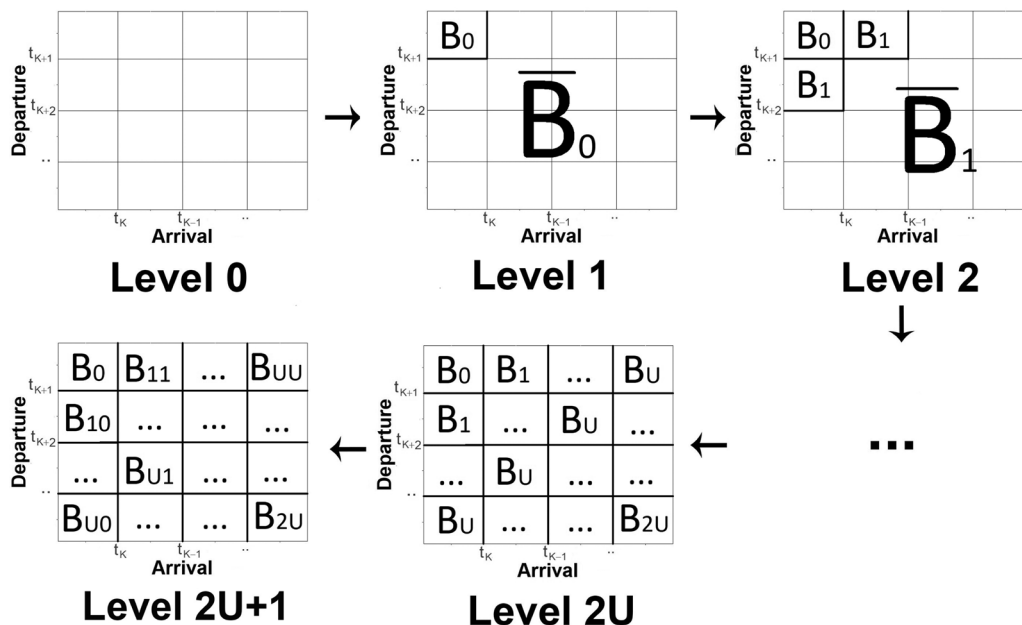
FIGURE 4 Partition of the PT in the case of ring-recovery data for the matrix $\mathbf{n}^k$ corresponding to the individuals marked at time t_k. The sample space is split into the set of individuals staying for less than one sampling occasion, B_0, and the set of individuals staying for more than one sampling occasion, B_1. After this process is repeated for $2U$ sampling occasions, each set is split according to the entry intervals.

if $\phi_{j,a}$ is the probability of an individual of age a that is present at time t_j to remain until the following sampling occasion, that $\phi_{j,a} \equiv \phi_a$. Because of this assumption, the posterior distributions on the survival probabilities are the same for each time of marking.

Assuming also a Beta prior for the recovery probability λ, the model can be summarized as

$$R_{k,k+j} \sim \text{Binomial}\left(\sum_{f=0}^{U} n_{fj}^k, \lambda\right), n_{fl}^k \sim \text{Multinomial}(m_k, \omega_{fl})$$

$$\{\omega_{fl}\} \sim \text{PT}(\Pi, \alpha), \lambda \sim \text{Beta}(a_0, b_0). \tag{3}$$

We show an application of this model in Section 6.2.

6 | CASE STUDIES

## 6.1 | Capture–recapture and count data

We jointly apply the models of Section 5.1 and 5.2 to a count and a CR dataset, respectively, collected on a population of Eurasian Spoonbills (*Platalea leucorodia*) in the southern Po delta, in North-East Italy. Birds are captured as chicks in previous years and are uniquely marked. The CR dataset is collected by resighting adult birds through photographs obtained using camera traps and by visiting their nests on eight separate sampling occasions. No attempt is made to mark new adult individuals and instead only previously marked individuals can be resighted. In addition, unmarked birds are detected on each sampling occasion. In this case, since there are fewer than 40 resighted individuals, we perform inference on the marked individuals resighted by modeling explicitly their individual entry and exit intervals, as mentioned in Section 5.2.

As opposed to the model described in Section 5.2, we have two different populations of individuals, the ones already marked and the unmarked, and hence, we have to modify the model that has been introduced before. We model the marked individuals resighted by introducing the variables (t_1^i, t_2^i), where t_1^i and t_2^i are the entry and exit interval of the ith marked individual resighted, respectively, whereas the marked individuals never resighted and the unmarked individuals are summarized in two matrices $\{n_{fl}^M\}$ and $\{n_{fl}^U\}$, respectively, where similarly to the count data model of Section 5.1, cell (f, l), corresponds to the number of individuals first available for capture at time t_f and last available at time t_l. We assume the same distribution on the entry and exit intervals of marked and unmarked birds.

The number of marked birds, N_M, and the number of unmarked birds, N_U, are assigned two Poisson prior distributions, $N_M \sim \text{Poisson}(\omega_M)$ and $N_U \sim \text{Poisson}(\omega_U)$, respectively. The model for n_{fl}^M and n_{fl}^U can therefore be summarized as $\begin{cases} n_{fl}^M \sim \text{Poisson}(\omega_M \times \omega_{fl}) \\ n_{fl}^U \sim \text{Poisson}(\omega_U \times \omega_{fl}) \end{cases}$. We assume that each marked individual can be resighted with probability p_R by visiting its nest and through camera traps with
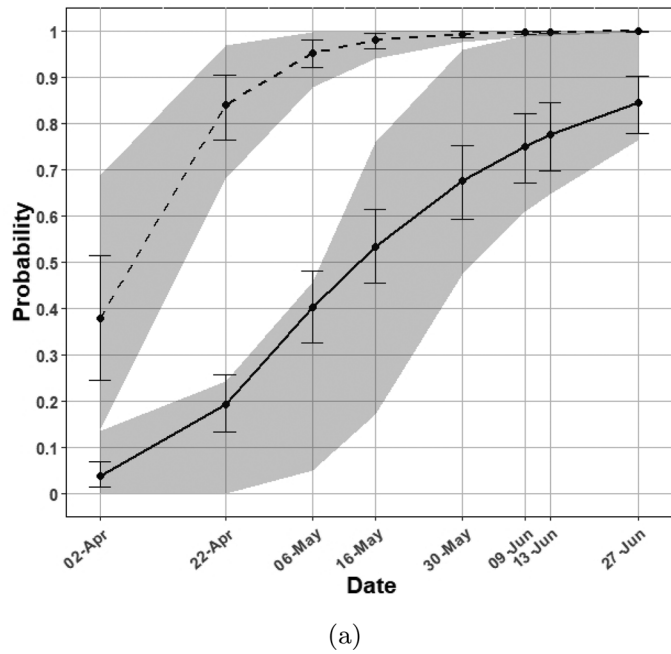
	2.5%	Mean	97.5%
N_M	39	48	59
N_U	45	65	91
p_R	.30	.38	.48
p_C	.21	.29	.38

(b)

FIGURE 5 (A) Posterior mean of the cumulative distribution function of entry intervals (dashed line) and exit intervals (solid line), with confidence bars showing the 95% posterior credible intervals. The shaded areas show the central 95% probability interval of the prior distributions. The ticks on the *x*-axis represent the sampling occasions. The ticks on the *x*-axis represent the sampling occasions. (B) Posterior summaries of population sizes of marked (N_M) and unmarked (N_U) individuals, and resighting probabilities in the nest (p_C) and using camera traps (p_R)

probability p_C. As the unmarked birds are also detected through camera traps, we assume that they can be detected with the same probability p_R of resighting an already marked bird. We note that this is a realistic, but also necessary, assumption in this case, as otherwise the counts do not include enough information to separately estimate the population size N_U of unmarked birds and the probability of resighting an unmarked bird, p_R, if that is different to marked birds. In Diana et al. (2018), this identifiability issue was overcome by employing informative prior distributions.

We choose a uniform prior distribution for the resighting probabilities and a Gamma prior distribution with mean 50 and variance 4000 for the intensities ω_M and ω_U of the two population sizes. The entry and exit distributions are centered on Laplace distributions $\text{Lap}(\mu_1, \lambda_1)$ and $\text{Lap}(\mu_2, \lambda_2)$ respectively, where $\mu_1 \sim \text{N}(t_1, 30^2)$, $\mu_2 \sim \text{N}(t_4, 30^2)$ and $\lambda_i \sim \text{Gamma}(a_\lambda, b_\lambda)$, such that $\mathbb{E}[\lambda_i] = 20$ and $\text{Var}[\lambda_i] = 80$, where days are the unit of measure. The resulting prior distributions on the entry and exit distributions are shown in Figure 5(A). Inference is performed using an MCMC algorithm by sampling from the posterior distribution of the individual entry and exit intervals (t_1^i, t_2^i), the two matrices $\{n_{fl}^M\}$ and $\{n_{fl}^U\}$, as well as the probabilities ω_{ij} of the PT and the resighting probabilities. The Markov chain Monte Carlo (MCMC) scheme is described in Section 6.1 of the Supporting Information.

The posterior means of the entry and exit cumulative distribution functions (CDFs) are shown in Figure 5(A). The CDF of the exit distribution suggests that individuals exit the area throughout the study period and, since it does not reach 1 at the end of the study period, that some birds remain at the site using the colony as a roost. Similarly, the CDF of the entry intervals is at around 0.4 at the start of the study, suggesting that a large proportion of birds are present at the site when sampling starts, with over 80% of birds estimated to have arrived by the second sampling occasion toward the end of April.

The posterior distributions of the two population sizes and resighting probabilities are shown in Figure 5(B). The camera trap resight probability, p_C, is slightly lower than the resight probability through nest visits, p_R, because cameras are pointed toward the nest, where it is unlikely to see floaters and prospectors. The difference between the population sizes of marked and unmarked birds is due to the fact that not all the chicks are marked each year, and out of those marked, only a small proportion return to breed as adults. In fact, local recruitment rate is thought to be around 0.12, whereas the proportion of immigrants on total number of recruits ranges from 0.49 to 0.83 (Tenan et al., 2017). Further details regarding the MCMC runs are given in the Supporting Information.

6.2 | Ring recovery

We apply the model for RR defined in Section 5.3 to a dataset collected in Minnesota, USA. A total of $100,127$ female mallards (*Anas platyrhynchos*) were banded throughout the course of the study, which lasted 51 years. Marking occurs from July to September, while recoveries occur during the hunting season immediately following marking, from September to January. Newly caught individuals can be recognized as juveniles if their age is less than 1 year at the time of capture and as adults otherwise. Therefore, individuals caught as juveniles are of known age at year of death. The entry and exit of the individuals correspond in this case to births and deaths, whereas the sampling occasions correspond to years.

We summarize the data in two $K \times K$ matrices, R^A and R^J, for adults and juveniles, respectively. The total number of juveniles and adults marked in sampling occasion k, are denoted by m_k^J and m_k^A, respectively. The first row of the matrix $\mathbf{n}^k$, $\{n_{0l}^k\}_{l=0,\dots,U}$, consists of the juveniles, and its sum is thus equal to m_k^J, whereas the rows from 2 to $U+1$ consist of the individuals marked as adults, and their sum is equal to m_k^A. We center the PT over a Weibull distribution $W(\lambda, k)$, with additional Gamma priors on λ and k. Assuming a Beta prior for the recovery probability, λ, the model can be summarized as

$$R_{k,k+l}^J \sim \text{Binomial}\left(n_{0l}^k, \lambda\right), \quad R_{k,k+l}^A \sim \text{Binomial}\left(\sum_{f=1}^{U} n_{fl}^k, \lambda\right),$$

$$n_{0l}^k \sim \text{Multinomial}\left(m_k^J, \omega_{0l}\right), \quad \{n_{fl}^k\}_{f>0} \sim \text{Multinomial}\left(m_k^A, \omega_{fl}\right),$$

$$\{\omega_{fl}\} \sim \text{PT}(\Pi, \{\alpha_{\lambda_W, k_W}\}), \quad \lambda \sim \text{Beta}(a_0, b_0),$$

$$\lambda_W \sim \text{N}\left(\lambda_W^0, \sigma_{\lambda_W}^2\right), \quad k_W \sim \text{N}\left(k_W^0, \sigma_{k_W}^2\right). \tag{4}$$

The first row of matrix n_{fl}^k contributes to the recoveries R^y for the juveniles and the rest of the matrix contributes to the recoveries R^a of the adults.

We perform inference by updating the random matrices n_{fl}^k, the probabilities ω_{fl} and the recovery probability λ. Details of the MCMC sampler are presented in Section 6.2 of the Supporting Information.

We assume $\lambda_W^0 = k_W^0 = 1$ and $\sigma_{\lambda_W}^2 = \sigma_{k_W}^2 = 0.5$. This prior reflects an uninformative prior distribution on the hazard, as it can be seen in Figure S6 of the Supporting Information. Moreover, we assume a Beta(1,1) prior distribution for the recovery probability λ and we set the upper bound U on the age of the individuals to 18, which is a conservative upper bound because no individual was recovered after 14 years of being marked. We compare this approach with a nonparametric unconstrained approach, where uniform prior distributions are assigned on the
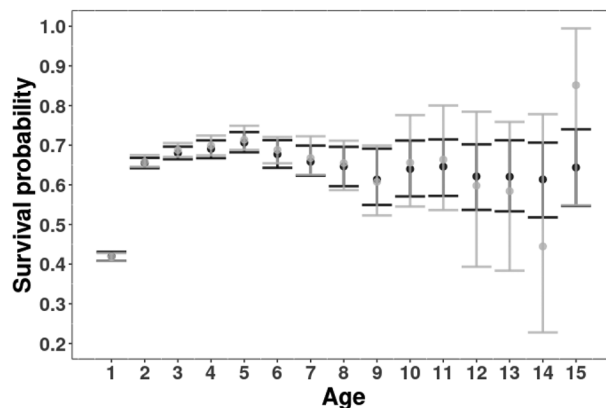

FIGURE 6 Case study on mallards. 95% posterior credible intervals of the probabilities ϕ_j of surviving to age $j+1$ conditional on being alive at age j for the unconstrained PT model (gray) and PT with Weibull centering distribution (black)

masses in each split of the PT (which is equivalent to $\alpha_{\epsilon_1,\dots,\epsilon_m} = 1$).

The posterior mean of the hazards is presented in Figure 6. The overall trend agrees with the general pattern observed for bird populations, with survival being lower in very young and older ages. A similar pattern was observed by McCrea et al. (2013) when analyzing RR data for mallards (*Anas platyrhynchos*) and by Jimenez-Muñoz et al. (2019) for blackbirds (*Turdus merula*), but we note here that we have not employed a parametric curve to enforce this pattern, and instead it has been completely driven by the data. We also show the results obtained using the unconstrained approach described above. As can be seen, the estimates for older ages for the model using the Weibull centering distribution are smoother, with narrower credible intervals and avoid boundaries, typically seen as a result of data sparseness at older ages, a result with further highlights the benefit of a BNP model centered on a parametric distribution. Finally, the posterior mean of the recovery probability is equal to 0.123 (95% PCI: 0.121–0.125), which is in line with findings of similar studies (McCrea et al., 2012). Further details regarding the MCMC runs are given in the Supporting Information.

7 | CONCLUSION

We introduced a framework for defining models for ecological data on open populations based on the PT prior. The advantage of this framework is that a wide variety of assumptions can be made on the model structure by changing the tree structure, as a result of the flexibility of the PT. We have applied our framework to different types of commonly collected ecological data, such as CR, RR, and count data. We have also introduced the RPT, which allows us to

place constraints on the model parameters, such as age-dependent survival probabilities or time-dependent exit probabilities. This assumption follows from the use of specific RPT structures that imply posterior independence of the entry and exit or LOS distributions, and can be relaxed by removing the RPT structure, which, however, results in a large number of unrestricted parameters to estimate.

It is easy to extend the model to other protocols, as, for example, removal data (Matechou et al., 2016; Otis et al., 1978). Removal data are collected by repeatedly visiting the site and removing all caught individuals. In this case, an approach similar to CR and RR can be employed, with the time of exit known for the removed individuals and unknown number of individuals not removed by the end of the study.

The RPT, which assumes that RVs are the same across different trees, represents the strongest form of sharing as it assumes complete pooling of the RV across the different trees. This implies particular forms of association between entry and exit or entry and LOS. Alternatively, it is possible to assume intermediate forms of sharing, which are between complete pooling and no sharing at all, by considering a hierarchical prior among the different trees. This form of sharing can be achieved by using a Hierarchical PT (Christensen & Ma, 2020) and defining the model by centering the PT in each group over a common PT prior. An alternative definition is to use a logistic PT (Jara & Hanson, 2011), which makes use of the logistic normal in place of the Beta distributions. These models make weaker assumptions and so have the potential to learn more about the association between entry and exit or entry and LOS. The use of the logistic normal PT gives interesting hints for further extensions. For example, by taking advantage of the normal structure, it is possible to define regression models or time-series models, with the possibility of using the Polya-Gamma scheme (Polson et al., 2013) for a conjugate scheme for inference.

In our modeling framework, different ecological assumptions correspond to different trees or changes in the dependence structure of the RVs on the tree. The question of evaluating evidence in favor of different assumptions can be addressed using standard Bayesian model selection methods. For example, model selection can be performed by evaluating the evidence of each model by using Bayes factors. Alternatively, as a change in the ecological assumption entails a change in the structure of the trees, model selection can be performed by employing an additional prior on these different tree structures and computing the posterior of the structure of trees as part of the parameter space.

Our modeling framework also provides an alternative approach to inferring population size, compared to, for example, N-mixture models for count data (Royle, 2004)

and data-augmentation approaches for CR and related data (Royle & Dorazio, 2012), which does not rely on specifying an upper bound for the population size and, as it does not infer individual entry and exit times, is also efficient for large populations.

DATA AVAILABILITY STATEMENT
The data that support the findings in this paper are available from the corresponding author upon reasonable request.

ORCID
Alex Diana https://orcid.org/0000-0002-8130-2988
Eleni Matechou https://orcid.org/0000-0003-3626-844X
Simone Tenan https://orcid.org/0000-0001-5055-9193
Stefano Volponi https://orcid.org/0000-0002-7471-2268

REFERENCES
Christensen, J. & Ma, L. (2020) A Bayesian hierarchical model for related densities by using Polya trees. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 82(1), 127–153.

Dennis, E.B., Morgan, B.J., Brereton, T.M., Roy, D.B. & Fox, R. (2017) Using citizen science butterfly counts to predict species population trends. *Conservation Biology*, 31(6), 1350–1361.

Diana, A., Griffin, J. & Matechou, E. (2018) A Polya tree based model for unmarked individuals in an open wildlife population. In: *International Conference on Bayesian Statistics in Action*. Springer, pp. 3–11.

Diana, A., Matechou, E., Griffin, J. & Johnston, A. (2020) A hierarchical dependent Dirichlet process prior for modelling bird migration patterns in the UK. *The Annals of Applied Statistics*, 14(1), 473–493.

Ferguson, T.S. (1974) Prior distributions on spaces of probability measures. *The Annals of Statistics*, 2(4), 615–629.

Ford, J.H., Patterson, T.A. & Bravington, M.V. (2015) Modelling latent individual heterogeneity in mark-recapture data with Dirichlet process priors. *arXiv preprint arXiv:1511.07103*.

Jara, A. & Hanson, T.E. (2011) A class of mixtures of dependent tail-free processes. *Biometrika*, 98(3), 553–566.

Jimenez-Muñoz, M., Cole, D.J. et al. (2019) Estimating age-dependent survival from age-aggregated ringing data. *Ecology and Evolution*, 9(2), 769–779.

Lavine, M. (1994) More aspects of Polya tree distributions for statistical modelling. *The Annals of Statistics*, 22(3), 1161–1176.

Lo, A.Y. (1984) On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics*, 12(1), 351–357.

Lyons, J.E., Kendall, W.L., Royle, J.A., Converse, S.J., Andres, B.A. & Buchanan, J.B. (2016) Population size and stopover duration

estimation using mark–resight data and Bayesian analysis of a superpopulation model. *Biometrics*, 72(1), 262–271.

Matechou, E. & Caron, F. (2017) Modelling individual migration patterns using a Bayesian nonparametric approach for capture–recapture data. *The Annals of Applied Statistics*, 11(1), 21–40.

Matechou, E., Dennis, E.B., Freeman, S.N. & Brereton, T. (2014) Monitoring abundance and phenology in (multivoltine) butterfly species: a novel mixture model. *Journal of Applied Ecology*, 51(3), 766–775.

Matechou, E., McCrea, R.S., Morgan, B.J., Nash, D.J. & Griffiths, R.A. (2016) Open models for removal data. *The Annals of Applied Statistics*, 10(3), 1572–1589.

Matechou, E., Morgan, B.J., Pledger, S., Collazo, J. & Lyons, J. (2013) Integrated analysis of capture–recapture–resighting data and counts of unmarked birds at stop-over sites. *Journal of Agricultural, Biological, and Environmental Statistics*, 18(1), 120–135.

McCrea, R.S., Morgan, B.J., Brown, D.I. & Robinson, R.A. (2012) Conditional modelling of ring-recovery data. *Methods in Ecology and Evolution*, 3(5), 823–831.

McCrea, R.S., Morgan, B.J.T. & Cole, D.J. (2013) Age-dependent mixture models for recovery data on animals marked at unknown age. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(1), 101–113.

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

Pledger, S., Efford, M., Pollock, K., Collazo, J. & Lyons, J. (2009) Stopover duration analysis with departure probability dependent on unknown time since arrival. In: Thomson, D.L., Cooch, E.G. & Conroy, M.J. (Eds.) *Modeling demographic processes in marked populations*. Springer, New York, pp. 349–363.

Polson, N.G., Scott, J.G. & Windle, J. (2013) Bayesian inference for logistic models using Pólya–Gamma latent variables. *Journal of the American Statistical Association*, 108(504), 1339–1349.

Royle, J.A. (2004) N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, 60(1), 108–115.

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.

Tenan, S., Fasola, M., Volponi, S. & Tavecchia, G. (2017) Conspecific and not performance-based attraction on immigrants drives colony growth in a waterbird. *Journal of Animal Ecology*, 86(5), 1074–1081.

SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 2, 4, 5 and 6, along with the code, are available with this paper at the Biometrics website on Wiley Online Library.

How to cite this article: Diana, A., Matechou, E., Griffin, J., Arnold, T., Tenan, S. & Volponi, S. (2022) A general modeling framework for open wildlife populations based on the Polya tree prior. *Biometrics*, 1–13. https://doi.org/10.1111/biom.13756