Biometrical Journal

Finite mixtures in capture–recapture surveys for modeling residency patterns in marine wildlife populations

Gianmarco Caruso[1,2] ⓘ | **Pierfrancesco Alaimo Di Loro**[3] ⓘ | **Marco Mingione**[4] ⓘ |
Luca Tardella[2] ⓘ | **Daniela Silvia Pace**[5,6] ⓘ | **Giovanna Jona Lasinio**[2] ⓘ

[1]MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

[2]Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy

[3]Department GEPLI, Libera Univerità Maria Ss. Assunta (LUMSA), Rome, Lazio, Italy

[4]Department of Political Sciences, University of Roma Tre, Rome, Italy

[5]Department of of Environmental Biology, Sapienza University of Rome, Rome, Italy

[6]Institute for the Study of Anthropogenic Impacts and Sustainability in the Marine Environment, CNR, Trapani, Italy

Correspondence
Gianmarco Caruso, MRC Biostatistics Unit, University of Cambridge, Cambridge, UK.
Email: gianmarco.caruso@mrc-bsu.cam.ac.uk

Abstract

This work aims to show how prior knowledge about the structure of a heterogeneous animal population can be leveraged to improve the abundance estimation from capture–recapture survey data. We combine the Open Jolly-Seber model with finite mixtures and propose a parsimonious specification tailored to the residency patterns of the common bottlenose dolphin. We employ a Bayesian framework for our inference, discussing the appropriate choice of priors to mitigate label-switching and nonidentifiability issues, commonly associated with finite mixture models. We conduct a series of simulation experiments to illustrate the competitive advantage of our proposal over less specific alternatives. The proposed approach is applied to data collected on the common bottlenose dolphin population inhabiting the Tiber River estuary (Mediterranean Sea). Our results provide novel insights into this population's size and structure, shedding light on some of the ecological processes governing its dynamics.

KEYWORDS
abundance estimation, Bayesian modeling, capture–recapture, finite mixtures

1 | INTRODUCTION

Capture–Recapture (CR) methods are statistical techniques widely employed to estimate the size of an elusive population for which it is impossible to get a complete enumeration. This task, applied initially to ecology for the study of fish

and wildlife populations (Matechou & Argiento, 2023; Otis et al., 1978; Wu & Holan, 2017), is now common to many other application fields such as epidemiology (Böhning et al., 2020; Chao et al., 2001; Maruotti et al., 2023) and social sciences (Böhning et al., 2018; Böhning & van der Heijden, 2009; Brittain & Böhning, 2009; Di Cecco et al., 2020; Farcomeni, 2022; Silverman, 2020). The term *capture* is inherited from the traditional way wild animals have been identified for decades—namely, through capture, marking, and release—but it is not necessarily intended for its physical sense anymore. Researchers increasingly adopt noninvasive methods for monitoring wild populations to minimize the costs and impact on the population of interest. Among those, photo-identification (Pace et al., 2021; Royle et al., 2009) and DNA sampling (Bravington et al., 2016; Morin et al., 2016) are becoming increasingly popular as they minimize behavioral responses that may bias the final estimates (see Alunni Fegatelli & Tardella, 2013, and references therein).

Original applications of such methods date back to the beginning of the 20th century and were based on standard homogeneity assumptions on the population structure and the identification process (Amstrup et al., 2010; Le Cren, 1965). The literature is now rich in alternatives that can address a large variety of deviations from such basic model assumptions and suit situations where, for example, individuals exhibit heterogeneous behaviors (Pledger, 2000), sampling occurs in continuous time (Altieri et al., 2023), stop-over sites are present (Matechou et al., 2013; Worthington et al., 2019; Wu et al., 2021), temporary emigration is allowed (Zhou et al., 2019), and so on. For an exhaustive review, see King and McCrea (2019) and references therein.

Our work concentrates on the abundance estimation of a *common bottlenose dolphin* population inhabiting a delimited area over multiple years. Such population is known to be open and calls for using the Open Jolly-Seber (JS) model, a standard CR framework for open populations (Amstrup et al., 2010). Furthermore, the established ecological literature affirms that it comprises individuals with different residency patterns and results in a population clustered into groups with different levels of site fidelity. Hence, the homogeneity assumption of the standard CR framework cannot hold, and the behavior of individuals belonging to different groups (i.e., entrance, capture, and survival) must be described by different parameters. Accurate estimation of the clustering structure and parameters is of utmost interest to ecologists who want to describe the population's dynamics and inform conservation policies. A widespread practice in such a heterogeneous setting is to consider the inclusion of individual covariates that can help explain the differences among the population's members. However, informative covariates are often unavailable and the heterogeneity is entirely latent, as it is in the case study under consideration.

Finite mixture models (FMMs) represent the natural solution to this impasse. Each individual is assigned to a different mixture component with its own set of common and distinct parameter values. FMM approaches to CR models have been successfully employed in a likelihood-based framework both in closed population (Dorazio & Royle, 2003; Pledger, 2000, 2005) and open population (Guéry et al., 2017; Pledger et al., 2010) settings. From the Bayesian perspective, attempts have been made to model heterogeneity in detection and behavioral effects in closed populations by Ghosh and Norris (2005). More recently, Turek et al. (2021) proposed a nonparametric FMM with an unknown number of components and different capture probabilities.

We build on Pledger et al. (2003, 2010)'s FMM extension to the Open JS model to account for latent heterogeneity. We embed this finite mixture approach in Royle and Dorazio (2008, 2012)'s parameter-expanded data-augmentation (PX-DA) framework, which turns out to be particularly convenient to fit Bayesian CR models via standard Markov chain Monte Carlo (MCMC) algorithms. We discuss its implementation challenges and introduce suitable prior specifications that mitigate the label switching and non-identifiability issues of FMM. The model is tested through an extended simulation study and applied to photo-identification CR survey data of the common bottlenose dolphins (*tursiops truncatus*) population inhabiting the area of the Tiber River estuary in the Mediterranean Sea. In particular, we show how the prior scientific knowledge on the population of interest can be leveraged to specify a parsimonious FMM tailored to its supposed structure and check its validity. When this is the case, we show that it can sensibly improve the performances of more comprehensive specifications.

The remainder of the paper is organized as follows: Section 2 describes the motivation and intuitions behind our work, with a brief description of the data that will be considered later on; Section 3 illustrates a Bayesian hierarchical formulation of the Pledger et al. (2003, 2010)'s JS class of mixture models within the Royle and Dorazio (2008), Royle and Dorazio (2012)'s DA framework and introduces our modeling proposal as a parsimonious alternative; Section 4 reports a simulation experiment to investigate different aspects of the model and its estimation; Section 5 provides the model choice and results on the set of data originally introduced as the motivating example.
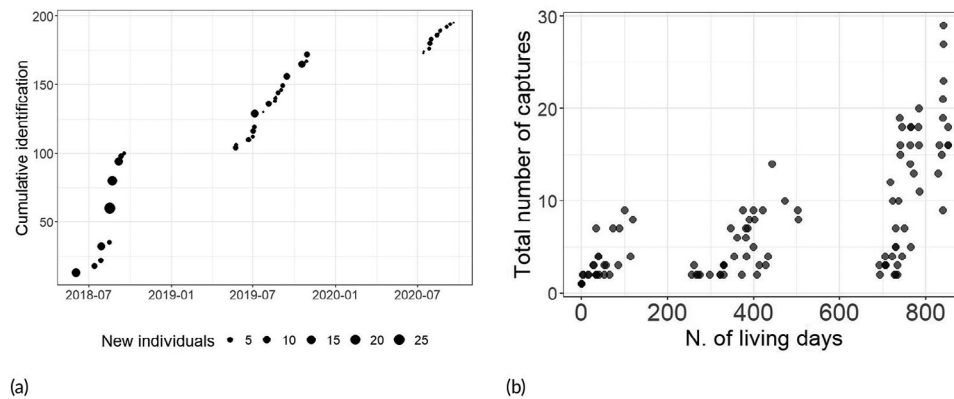
FIGURE 1 (a) Cumulative number of individual identifications with size proportional to the number of newly identified individuals. (b) Total number of captures by individual.

2 │ MOTIVATING EXAMPLE

This work is motivated by the need to get an abundance estimate of the population of the common bottlenose dolphins inhabiting the area of the Tiber River estuary in the Mediterranean Sea. Boat-based daily surveys have been conducted between 2018 and 2020 under favorable weather conditions to collect photographic and acoustic data of the specimens encountered in the study area during the search (Pace et al., 2022a, 2022b, 2022c; Papale et al., 2021). The photoidentification technique was used to identify unique individuals over multiple sampling occasions and to build single capture histories. For this analysis, we focus on the so-called *well-marked* individuals because the probability of their misidentification can be assumed to be negligible.[1] As a consequence, the final estimates are related to the subset of well-marked individuals only, representing a portion of the population visiting the study area. Figure 1(a) shows the cumulative number of identified individuals across the different sampling occasions, where the size of each point is proportional to the number of newly identified individuals. Its trend is known as the *discovery rate*, which is maximum during the first year (notice that 50% of new identifications occurred in 2018, with a maximum of 25 newly identified individuals registered in August) and slowly decreases over time. Further details about the study area, the data collection process, and the analysis are available in Pace et al. (2021).

Most of the recent literature about common bottlenose dolphins converges toward the identification of three groups characterized by different levels of site fidelity to a specific area: from the most to the least frequently present (Dinis et al., 2016; Haughey et al., 2020; Hunt et al., 2017; La Manna et al., 2022). This feature is of utter interest to biologists interested in disentangling the permanent or semipermanent population from the transient one. Generally speaking, one group is composed of individuals who (almost) never leave the study area; these are usually referred to as *resident* individuals, observable on many occasions and for long periods of time, and expected to have the largest number of captures (see Figure 1b). The other group includes individuals who are not continuously present in the study area but regularly visit it; these are called *part-time* individuals, observable throughout a wide time window but usually encountered at occasions far apart in time. The last group comprises individuals who enter the study area only once in their lifetime for a short time window and whose captures are rare; these are *transient* individuals who are observable only on occasions occurring on close dates.

This inherent heterogeneity cannot be neglected without hindering the reliability of the final abundance estimates (Gimenez et al., 2018). To this end, Pace et al. (2021) apply a hierarchical clustering and include the group labels in the POPAN-JS framework (Schwarz & Arnason, 1996) to model heterogeneity in the entrance, capture, and survival probabilities. They find signs of the expected structure and estimate significant differences in most components of the JS model. While practical for understanding the underlying structure of the population of interest, this two-step approach has some relevant issues. First of all, there is neither quantification nor propagation of the uncertainty of the classification step onto the modeling step; this can bias the final estimates and yield overconfident conclusions. Second, the same information

---

[1] Photoidentification is based on matching the same marks across different pictures; the level of the mark (poor, fair, well) is an index of how identifiable a dolphin is.

set is used twice in two different statistical procedures, where the latter is performed after conditioning on the first; this can lead to a confirmation bias in favor of the original hypothesis.

This paper proposes to unify the two steps into a joint statistical procedure. We embed FMM into the Open-JS framework in a Bayesian hierarchical setting, allowing for the estimation of cluster labels and population size altogether while properly propagating the uncertainty at all levels (Clark & Gelfand, 2006).

3 | STATE OF THE ART AND PROPOSED MODEL

Let D be the number of distinct individuals observed at least once during T distinct sampling occasions. Data are collected in a $(D \times T)$ matrix, say $\mathbf{Y}$, with the generic element recording whether individual $i = 1, \ldots, D$ individual has been observed ($y_{it} = 1$) or not ($y_{it} = 0$) at occasion $t = 1, \ldots, T$. Hence, the rows of $\mathbf{Y}$ contain the capture histories of all the encountered individuals. The open-population JS model (Schwarz & Arnason, 1996) envisions individuals entering (i.e., via birth or immigration) and exiting (i.e., via death or emigration) the population during the sampling occasions. The emigration is assumed to be permanent for identifiability purposes, that is, once an individual has left the population, it cannot return to it. Furthermore, the JS models assume that all captures are independent across individuals and over time. The latter is achieved by considering the existence of a superpopulation of unknown size, say N_{super}, accounting for all the individuals potentially available (encountered or not) in the study area between the first and the last sampling period. The superpopulation size is the main parameter of interest and determines the dimension of the model parameter space. Its typical estimation in the Bayesian framework involves jumping between spaces of different dimensions, hence requiring the implementation of Reversible Jump MCMC algorithms (Brooks et al., 2000). Here, we adopt the DA approach proposed by Royle and Dorazio (2008) to keep the dimension of the parameter space *fixed* throughout the iterations, thus bypassing the need for the Reversible Jump MCMC (Arnold et al., 2010).

3.1 | Parameter expanded data-augmentation formalization

The PX-DA approach by Royle and Dorazio (2008) sets the dimension of the parameter space equal to a fixed value $M \gg N_{super} \geq D$. In this way, the varying dimension issue is converted into a more manageable *missing data* problem in the multistate processes context (see Section A of the Appendix). The rows of the observed data matrix $\mathbf{Y}$ are augmented to M, hence defining $\mathbf{Y}_{\text{aug}} = \{\mathbf{Y}, \mathbf{0}_{M-D}\}$, where $\mathbf{0}_{M-D}$ is an $(M - D) \times T$ matrix of zeroes. M must be set such that $N_{super} \in \{D, D + 1, \ldots, M\}$, and consequently, $N_{super} - D$ among the $M - D$ rows of zeroes correspond to individuals who belong to the superpopulation but have never been encountered. The remaining $M - N_{super}$ correspond to *pseudo*-individuals who have never been part of the population during the observation window and hence do not belong to the superpopulation. The data generation process assumes that an individual can be recruited into the population at the beginning of each sampling period only if it has never been recruited on previous occasions. On the other hand, individuals who have already been recruited can leave the population between two subsequent sampling periods. All the D observed individuals will eventually be recruited in the population between the first and the last capture occasions, as they have been captured at least once. This dynamic is controlled through two time-varying latent binary variables: the first one, r_{it}, is equal to 1 if and only if individual i is recruitable at time t (0, otherwise); the second one, z_{it}, is equal to 1 if and only if individual i belongs to the population at time t (0, otherwise).

All M individuals are recruitable at the first occasion (i.e., $r_{i1} = 1$, $i = 1, \ldots, M$), whereas they become permanently nonrecruitable once they have entered the population. Let ρ_t, $t = 1, \ldots, T$ be the recruitment probabilities, that is, the probability that an available (not yet entered) individual in the augmented dataset is recruited in the population at time t. Let ϕ_t, $t = 2, \ldots, T$, be the apparent survival probability,[2] that is, the probability that a recruited individual is in the population at the following sampling occasion. The following rules govern the latent label process: $z_{i1} \sim Bern(\rho_1)$, $i = 1, \ldots, M$, $r_{it} = \min\{r_{i,t-1}, 1 - z_{i,t-1}\}$ and $z_{it} \mid z_{i,t-1}, r_{it} \sim Bern(\phi_t \cdot z_{i,t-1} + \rho_t \cdot r_{it})$, $t = 2, \ldots, T$. Thus, the distribution of the generic element of the augmented data matrix can be expressed conditionally on z_{it} as:

$$y_{it} \mid z_{it} \sim Bern(p_t \cdot z_{it}), \tag{1}$$

---

[2] The term *apparent* is included as permanent emigration and mortality are indistinguishable, and thus, treated as the same phenomenon in JS-type models.

where p_t is the capture probability at occasion t. Since $y_{it} = 0$ almost surely when $z_{it} = 0$, (1) is a zero-inflated binomial model. Notice that, given M, the marginal likelihood of the capture histories can be expressed as:

$$\mathcal{L}_M(\boldsymbol{y} \mid \theta) = \int_{\mathcal{Z}} \prod_{i=1}^{M} \prod_{t=1}^{T} p(y_{it} \mid z_{it}) \cdot p(z_{it} \mid z_{i,t-1}, r_{it}) \, p(dz_{it}), \tag{2}$$

where $\mathcal{Z} \subset \{0, 1\}^{M \times T}$ contains all possible z_{it} configurations and θ is the full set of model parameters $\theta = \{p_t, \phi_t, \rho_t\}_{t=1}^{T}$. The integral in the likelihood expression has no general closed-form solution and numerically marginalizing out the latent components can be computationally intensive. This problem is common to the hidden Markov models (HMMs) literature, in which some efficient techniques based on the forward algorithm have been proposed to efficiently solve the integration problem Worthington et al. (2019). Alternatively, MCMC algorithms provide a viable solution to work, iteration after iteration, with the conditional likelihood:

$$\mathcal{L}_M^c(\boldsymbol{y} \mid \theta, \boldsymbol{z}) = \prod_{i=1}^{M} \prod_{t=1}^{T} p(y_{it} \mid z_{it}),$$

and approximate the posterior distribution of all the quantities of interests in a Bayesian setting that provides full uncertainty quantification. Furthermore, standard MCMC methods are now relatively easy to implement by practitioners due to the availability of software like JAGS (Plummer, 2003) or NIMBLE (de Valpine et al., 2017).

Marginally, the hierarchical model specification resulting from Equation (1) implies that $N_{super} \sim Binom(M, \psi)$, where ψ is the overall inclusion probability (throughout all occasions) of any individual in the superpopulation. Royle and Dorazio (2008) show that ψ is linked to the recruitment probabilities through the following equation:

$$\psi = 1 - \prod_{t=1}^{T} (1 - \rho_t). \tag{3}$$

In particular, (3) implies that $\mathbb{E}[N_{super} \mid M, \rho_1, \dots, \rho_T] = M\left[1 - \prod_{t=1}^{T} (1 - \rho_t)\right] = M\psi$. Hence, choosing the prior distribution for $\rho_1, \dots, \rho_T$ is crucial to determining the prior on N_{super}. Dorazio (2020) demonstrates that the prior:

$$\rho_t \sim Beta\left(\frac{1}{T}, 2 - \frac{t}{T}\right), \quad t = 1, \dots, T, \tag{4}$$

induces an objective prior on N_{super}. In terms of practical inference, the estimated population size at each time t and the overall superpopulation size can be derived through the latent variables z's, namely, $N_t = \sum_{i=1}^{M} z_{it}$ and $N_{super} = \sum_{i=1}^{M} \mathbb{1}_{\{\sum_{t=1}^{T} z_{it} > 0\}}$.

3.2 | Jolly–Seber finite mixture modeling for open populations

Most populations are composed of individuals with heterogeneous behaviors. In some cases, the heterogeneity can be assumed to be well described by a finite number (say G) of different patterns. These depend on the individual's latent traits, and their consideration requires complex modeling tools, such as FMM. The underlying assumption of FMM is that each unit can belong to only one group $g = 1, \dots, G$, with unknown prior probabilities w_g ($\sum_g w_g = 1$). The individuals belonging to different groups in open population CR studies may have different capture, recruitment, or survival parameters. In the most general specification, the relative order among the parameters of different groups can change at each time t (e.g., group one could have the highest detection rate at the first sampling occasion but the lowest at the second one). This JS-type mixture model is known as the interactive heterogeneous model (IHM) and its very rich specification depends on too many parameters for successful model fitting (see Pledger et al., 2010, for further discussion). Pledger et al. (2003, 2010) explore simpler specifications that could adequately represent the population structure and introduce a convenient notation to navigate through all possible submodels. Let t and h be the time and group heterogeneity effects, respectively. Different expressions correspond to different modeling structures: constant in time and homogeneous across

groups $(\cdot)$; time-varying but homogeneous across groups (t); constant in time but heterogeneous across groups (h); time-varying and heterogeneous across groups with separable interaction $(t + h)$; and time-varying and heterogeneous across group with nonseparable interaction $(t \times h)$. For example, the IHM corresponds to $\{[\rho_{t \times h}, \phi_{t \times h}, p_{t \times h}]_G\}$, where the underlying population is supposed to be composed by G classes. If we want to specify a model whose heterogeneous group effect lies in the capture probabilities only, we write $\{[\rho_t, \phi_t, p_{t \times h_G}]\}$. The subscript is moved to highlight that the mixture of G components is related only to detection. We will take advantage of this notation in the sequel of the paper.

3.3 | Modeling class heterogeneity using finite mixtures within the PX-DA approach

We embed the PX-DA formalization of the open JS model into FMM by adding one layer of hierarchy in the original hierarchical specification.

Now, let $c_i \in \{1, \dots, G\}$ be the latent membership label of each individual $i = 1, \dots, M$ in $\mathbf{Y_{aug}}$. The full hierarchical specification is as follows:

$$
\begin{aligned}
y_{it} \mid z_{it}, c_i = g &\sim Bern(p_{gt} \cdot z_{it}), \\
z_{it} \mid z_{i,t-1}, r_{it}, c_i = g &\sim Bern(\phi_{gt} \cdot z_{i,t-1} + \rho_{gt} \cdot r_{it}), \quad r_{it} = \min\{r_{i,t-1}, 1 - z_{i,t-1}\}, \\
p_{gt} &\sim \pi_{p_g}(\cdot), \quad \phi_{gt} \sim \pi_{\phi_g}(\cdot), \quad \rho_{gt} \sim \pi_{\rho_g}(\cdot), \quad c_i \sim Multinom(1, (w_1, \dots, w_G)), \\
(w_1, \dots, w_G) &\sim \pi_{\mathbf{w}}(\cdot)
\end{aligned} \tag{5}
$$

where $\pi_{\cdot}(\cdot)$ refers to a generic prior distribution for the parameter. The hierarchical formulation of (5) further complicates the likelihood expression of Equation (2) to:

$$
\mathcal{L}_M(\mathbf{y} \mid \tilde{\theta}) = \sum_{g=1}^{G} w_g \cdot \int_{\mathcal{Z}} \prod_{i=1}^{M} \prod_{t=1}^{T} p_g(y_{it} \mid z_{it}) \cdot p_g(z_{it} \mid z_{i,t-1}, r_{it}) dz_{it},,
$$

where $\tilde{\theta} = \{\mathbf{w}, \{\theta_g\}_{g=1}^{G}\}$ includes the prior weights of the cluster components and the component-specific sets of parameters. Its evaluation needs further marginalization with respect to the latent group labels, and, once again, MCMC methods are a viable solution to achieve Bayesian estimation of all quantities of interest.

This modeling framework includes many possible specifications, according to what varies with time and across groups. We further generalize the model specification by adapting the survival mechanism to describe not equally spaced capture occasions. Indeed, the assumption of constant survival across identical time scales does not transfer to the not equally spaced scenario. When this is the case, that is, $l_t = (\tau_t - \tau_{t-1}), t = 2, \dots, T$ are the time differences between subsequent occasions—the survival probabilities should be appropriately compounded. Once the time scale is set (e.g., days, weeks, months, years, etc.), we have $\phi_{gt} = \phi_g^{l_t}$, where ϕ_g represents the survival probability across a single time unit on the chosen scale. In addition, following Pledger et al. (2003), a convenient and parsimonious way to express the time-varying capture probabilities is through the logit link, that is, $logit(p_{gt}) = \mu_g + \tau_t, t = 1, \dots, T$, where μ_g determines the overall average capture probability of each group and τ_t is an occasion-specific differential effect.

The last step of the Bayesian model specification involves the choice of prior distributions for all parameters and their hyperparameters. The natural prior for the mixture weights in FMMs is the Dirichlet distribution $Dir_G(\alpha_1, \dots, \alpha_G)$. It corresponds to a uniform distribution over the G-dimensional simplex when $\alpha_g = 1, \forall g$, x is usually adopted as weakly informative prior. For $\rho_{gt}, t = 1, \dots, T$, we follow Dorazio (2020) and use the prior in (4). General-purpose and weakly informative priors can be ascribed to μ_g. At the same time, we suggest considering a $N(0, \sigma^2)$ on each parameter τ_t, with σ^2 small, to induce only small time variations (in the logit-scale) on the capture probabilities. We further impose that $\sum_t \tau_t = 0$ to favor the otherwise weak identifiability of the μ_g and the τ's (Pledger et al., 2003). Nevertheless, such choices do not defend the model from the label-switching problem of the class-specific parameters typical of the FMM framework because of the likelihood invariance under permutations of the components' labels.

Choosing priors on class-specific parameters

The choice of the prior distribution on component-specific parameters of FMMs can be seen as a challenge and an opportunity. This class of models suffers from several sources of nonidentifiability. In our setting, the most affected are recruitment and survival probabilities. However, the final inference on the population size is usually robust with respect to their proper identification (see Mena & Walker, 2015, and references therein). The recruitment probabilities ρ_{gt} are model devices that allow the entrance of new individuals in the population, accounting for the openness of the population (Royle & Dorazio, 2008). They can be treated as nuisance parameters without a solid biological meaning and their identifiability is not of great concern. On the other hand, the capture and survival probabilities are of biological interest. They can be seen as indicators of how long and how often units of different types visits the sampling area.

A weakly informative or regularizing choice of their prior distributions can favor identifying these components. In turn, this can ease the identification of the membership labels already affected by the well-known *label-switching* problem. Among the different solutions that have been proposed in the literature, one relies on imposing an ordering constraint among the component-specific parameters, that is, $u_1 < \cdots < u_G$ (Chung et al., 2004; Diebolt & Robert, 1994; Richardson & Green, 1997). This task is trivial in the context of parameters belonging to the whole real domain, where truncating or shifting Gaussian distributions allows full control of the position and scale of the priors. The latter is a reasonable choice for the μ_g, $g = 1, \ldots, G$ parameters. However, when $u_g \in (0, 1)$, $\forall g$ as in the case of the survival probabilities, the conditional prior specification must account for their bounded domains. The standard solution would consider conditionally specified Uniform distributions. However, we exploit the alternative proposal of Alaimo Di Loro et al. (2022) that allows for more flexibility. It is based on the Beta distribution and its generalizations (*truncated* or *restricted*), and it is both effective in imposing the constraint and controlling for the shape and first moments of the induced marginal priors. In other words, one may set $u_1 \sim Beta(\alpha_1, \beta_1)$ and

$$u_g|u_{g-1} \sim tBeta(\alpha_g, \beta_g; u_{g-1}, 1), \quad \text{or} \quad u_g|u_{g-1} \sim rBeta(\alpha_g, \beta_g; u_{g-1}, 1) \qquad g = 2, \ldots, G, \tag{6}$$

where *tBeta* and *rBeta* represent the Truncated Beta and Restricted Beta, respectively, and it is possible to derive closed-form expressions for the marginal prior distribution of $u_g, g \geq 2$, or at least its first and second moments. Further computational details are given in Section B of the Appendix, along with formal proofs of the original results.

3.4 | The RPT model for the common bottlenose dolphin population

The population structure illustrated in Section 2 can be translated to reflect variations in the parameters of the JS-type PX-DA framework. We have that the resident (R) individuals, showing high site fidelity, should be characterized by high survival and capture probabilities; the part-time resident (P) individuals, with average site fidelity, should be characterized by high survival probabilities but eventually undetectable on some occasions; the transient individuals (T) only shortly visit the study area and hence should have very low survival. We want to build a suitably flexible encompassing model such that it is possible to establish the amount of evidence in favor of this assumption. In other words, we shall not enforce this exact structure in our model, but we can make it so that it is recognizable if present. For instance, we could allow the three groups to have group-specific survival and capture probabilities as in the full specification of (5). That might be a too flexible model with a huge number of parameters, therefore yielding highly uncertain estimates because of weak identifiability. Alternatively, we can propose a more parsimonious specification tailored to the supposed behavioral differences among individuals from different groups. In this way, some components can be assumed to be common to different groups, reducing the model's complexity and favoring its identifiability.

First of all, we model the population recruitment dynamic, assuming that each group is characterized by its own set of time-varying recruitability parameters, that is, $\rho_{R,t}$, $\rho_{P,t}$, $\rho_{T,t}$. This setting naturally induces a clustering in the relationship between the recruitment process and the superpopulation size N_{super}, yielding group-specific inflation parameters that express the total population as the sum of three subpopulations. This modifies the analytical expression of the expected superpopulation size as follows:

$$\mathbb{E}[N_{super}|\psi_R, \psi_P, \psi_T, w_R, w_P, w_T] = M \cdot \sum_{g=R,P,T} w_g \psi_g,$$

where w_g and ψ_g ($g = R, P, T$) are, respectively, the gth component-specific mixture weight and inflation parameter.

Second, we separate the three subpopulations into two distinct classes, characterized by different tendencies to stay in the study area. We can distinguish between short-term survivors, that is, transient individuals who visit the area for narrow time windows, and long-term survivors, that is, nontransient individuals (resident or part-time) who visit the area for wide time windows. To model such behavioral heterogeneity, we assume that one group (T, transient) has a smaller survival probability (ϕ_T) than the other two groups (ϕ_{NT}): $\phi_T < \phi_{NT}$. Notice that these two parameters represent the survival probabilities across a single time unit on the chosen scale. When the capture occasions are not equally spaced, the survival probabilities should be appropriately compounded along different lengths (cf. Section 3.3). In practice, given the two base survival probabilities ϕ_T and ϕ_{NT}, the actual survival probabilities at time t are:

$$\phi_{T,t} = \phi_T^{l_t}, \quad \phi_{NT,t} = \phi_{NT}^{l_t},$$

where $l_t = |t - t'|$, $t = 1, \dots, T$ are the time lags between the T subsequent occasions. Finally, the proposed model retains the time-varying structure of the detectability, as already introduced in Section 3.3, but discriminates between part-time (P) and non-part-time (NP) individuals introducing a *partial undetectability* component in the latter group. Loosely speaking, part-time individuals are allowed to be *undetectable* while *alive* (because temporarily not present in the study area) on some occasions chosen at random with probability $\delta \in (0, 1)$. This corresponds to modeling the capture probability on occasion t as:

$$p_{NP,t} = \text{logit}^{-1}(\mu + \tau_t) \text{ and } \quad p_{P,t} = (1 - \delta) \cdot p_{NP,t}, \quad t = 1, \dots, T,$$

for the NP and P individuals, respectively; of course, $p_{P,t} < p_{NP,t}$, $\forall t$. Notice that the larger δ is, the more the part-time group is separated from the resident group. This parameter plays a similar role to the completely random emigration parameter of Kendall et al. (1997) and it is needed to control the temporary emigration pattern of *part-time* individuals discussed in Section 2. Appendix C shows that such parametrization is indeed equivalent to temporary emigration under the simplifying assumption of emigration occurring at random. We name this model RPT as it encompasses three specific types of behavior. However, we would like to point out how the model does not enforce this interpretation. For instance, both survivals could be estimated to be high, or the undetectability parameter could be estimated as approximately equal to 1, and so on.

4 | SIMULATION EXPERIMENTS

We conduct a simulation experiment to assess the performances of the RPT model whenever it is well specified, that is, the data are generated according to the structure described in Section 3.4. We generate multiple sets of artificial data under alternative scenarios from the RPT with fixed parameters and then estimate a pool of models on them. Our main objective is twofold: (i) to evaluate the ability to recover the true values of the parameters, with a particular focus on N_{super}; (ii) to verify whether the RPT is chosen as the *best* among other alternatives based on some model selection criterion.

We consider four scenarios characterized by an increasing number of sampling occasions, that is, $T \in \{10, 20, 30, 40\}$, to verify the model performances for different time horizons. We suppose that in the first scenario (i.e., $T = 10$), all the captures are recorded within a relatively short period (e.g., within a year). Longer time horizons are included in the other scenarios, where a larger time gap (year gap) is assumed to occur every 10 occasions. Further details about the time lags are available in Section D of the Appendix. The month (and portion of months) is taken as the basic time unit to avoid the possible numerical instability related to the large values of the lags in terms of days. Note that this affects the interpretation of the survival probability parameter as it must be interpreted as the probability of surviving 1 month. We adopt the following parameters' values in all the scenarios. The survival probabilities are set to $\phi_T = 0.01$ and $\phi_{NT} = 0.997$. These two values may seem quite extreme at first glance. However, they guarantee that the short-time survivors (transient individuals) almost surely stay in the population for less than a year and that the long-term survivors (nontransient individuals) stay in the population for more than 3 months with a very high probability (> 0.99). Furthermore, notice that a monthly survival probability equal to 0.01 corresponds to a survival probability equal to 0.86 on a daily scale and equal to 0.34 on a weekly scale. On the other hand, a monthly survival probability equal to 0.997 corresponds to a probability of 0.87 on a 4-years scale. Therefore, the monthly scale appears as a good compromise to avoid a value that is too low for ϕ_T and a value that is too high for ϕ_{NT}. The capture probabilities are obtained by setting $\mu = 0$ and $\delta = 0.7$, and generating $\tau_t \sim N(0, 0.25)$ in each scenario. The recruitment parameter for transient individuals is fixed to $\rho_{T,t} = 0.02$ for all capture

occasions (assuming a negligible temporal variation for this group). At the same time, we consider the following structure for resident and part-time individuals:

$$\rho_{R,t} = \begin{cases} 0.4, & t = 1 \\ 0.02, & t = 10k+1, \; k = 1,2,3 \\ 0.0025, & \text{otherwise} \end{cases} \quad \text{and} \quad \rho_{P,t} = \begin{cases} 0.4, & t = 1 \\ 0.04, & t = 10k+1, \; k = 1,2,3 \\ 0.005, & \text{otherwise}. \end{cases}$$

The mixture weights for the three groups are set to $w_R = 0.2$, $w_P = 0.45$, and $w_T = 0.35$. We envision an augmented superpopulation of $M^* = 500$, which yields an expected superpopulation size $\mathbb{E}[N_{super}] \in \{170, 209, 243, 271\}$ for $T = 10, 20, 30, 40$, respectively. Notice that N_{super} increases with T as more individuals can visit the study area during a longer time horizon.

We simulate independent encounter histories for $K = 50$ pseudopopulations. Along with the RPT model, we consider 10 different alternatives in the class of JS-type models described in Sections 3.2 and 3.3, all having time-varying recruitment parameters. In the simplest case, we consider a model with homogeneous capture and survival probability (M_1: $\{\rho_t, \phi, p_t\}$). When the population is supposed to be structured in G groups, we suppose that each group has its own time-varying recruitment probability and that mixture components may vary by capture probability (M_2–M_4: $\{\rho_{t \times h_G}, \phi, p_{t+h_G}\}$, for $G = 2, 3, 4$), by survival probability (M_5–M_7: $\{\rho_{t \times h_G}, \phi_{h_G}, p_t\}$, for $G = 2, 3, 4$) or by both (M_8–M_{10}: $\{[\rho_{t \times h_G}, \phi_h, p_{t+h}]_G\}$, for $G = 2, 3, 4$). Each simulated dataset was augmented by 500 all-zero capture histories to implement the PX-DA approach so that $D_k + 500 = M_k \neq M^*$ for all simulated sets. We consider the prior setting described in Section 3. Notably, we specify an $N(0, 10)$ for the intercept μ. When the survival probability is the same for all the individuals, a standard Uniform prior is placed on that parameter. When two survival probabilities are considered (generically, ϕ_1 and ϕ_2), we use $\phi_1 \sim Beta(1, 2)$ and $\phi_2|\phi_T \sim tBeta(1, 1, \phi_T, 1)$, with the latter marginally yielding $\phi_2 \sim Beta(2, 1)$. Then enforcing $\phi_1 < \phi_2$ induces a slight repulsion between the two parameters. When one considers more than two survival parameters, constrained Uniform priors are instead chosen. These different prior choices for the models do not substantially affect the model selection criterion associated with each model but—conversely—are useful to induce a better separation between couples of survival parameters. Estimation is carried out using JAGS, in which we run two chains with 20,000 iterations each, discarding 5000 as burn-in and thinning by 2 the remainder to save storage space (Brooks et al., 2004).

We chose median posterior estimates instead of simple averages, mitigating the effect of anomalies that can result in occasionally low-informative datasets. In the same spirit, we rely on the *mean absolute error* (MAE) as an accuracy measure instead of the widely employed *root mean squared Error*. Interval estimation is assessed through the percentage of times the 95% credible intervals contain the true value of the parameter (i.e., the coverage) and the average 95% credible interval width (CIW). The overall goodness-of-fit is measured via the Watanabe–Akaike information criterion (WAIC, Watanabe & Opper, 2010), following the good practice of Gelman et al. (2014) whenever finite mixtures are fitted. We also report the overlapping index (OV, Pastore, 2018; Pastore & Calcagnì, 2019) between the posterior distributions of the two survival parameters, averaged over the $K = 50$ replicas. If $OV = 0$, the two distributions are completely separated, while if $OV = 1$, the two distributions perfectly overlap. This metric is particularly appealing to understand whether the posterior distributions of the group-specific parameters are well separated or not, justifying the related model parameterization. Finally, we investigate the fuzzy classification ability of the RPT model using multiclass area under the receiver operating characteristics Curve (mAUC) (Hand & Till, 2001). Let us remark that the classification performances can only be evaluated in the simulation setting and for the RPT model, as the true group labels are known and are consistent with the estimated ones. Notice that individuals who have not been observed are not provided with a capture history; thus, it is impossible to infer the group they belong to. Therefore, the estimated mixture weights (referred to as the full M_k-sized pseudopopulation) are not expected to align with the mixture weights employed in the data generation process. Indeed, the latter was used to generate the cluster labels of the whole pseudopopulation, of which some individuals (i.e., the pseudo-individuals) never become part of the real population.

Figure 2 shows the differences between the estimated ($\hat{N}_{super}$) and true (N_{super}) superpopulation size for each of the $K = 50$ replicas. The error is divided by the expected value of N_{super} in the corresponding scenario to allow for a meaningful comparison between scenarios having expected superpopulation sizes of different magnitudes. We notice that the RPT uniformly provides the best results overall. On the contrary, its competing models consistently underestimate or overestimate the superpopulation size. Notably, as T increases, the underestimation is more evident for those models that do not account for heterogeneity in survival probabilities (i.e., M_1–M_4)—cf. Pledger et al. (2003, 2010)—while the overestimation is substantial for those that consider an excessive number of parameters (i.e., M_8–M_{10}). Intuitively, a greater number
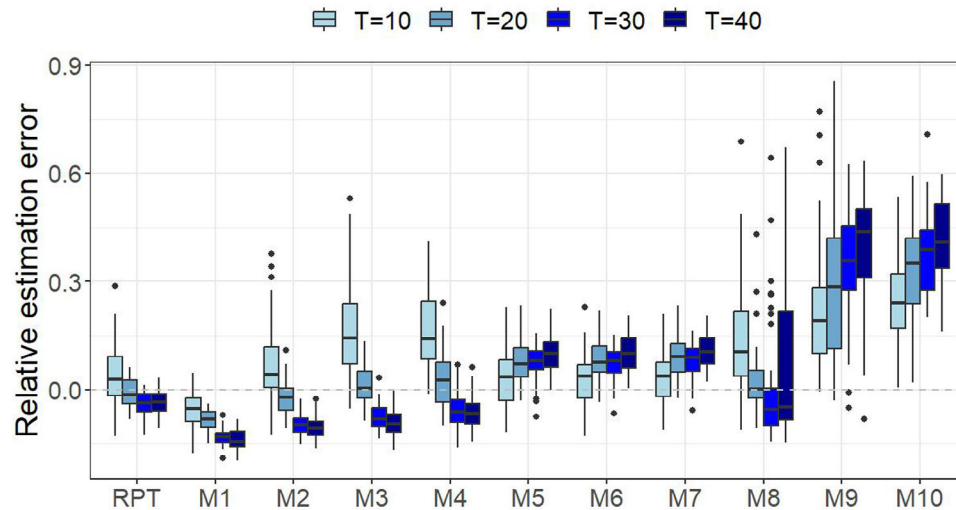
FIGURE 2   Relative estimation error of the superpopulation (N_{super}) abundance for increasing number of sampling occasions (T), calculated for each of the $K = 50$ independent replicas, by the RPT and the 10 alternative Pledger et al. (2010)'s mixture models.

of parameters controlling the capture probabilities tend to infer an excessive number of uncaught individuals from the zero histories.

Table 1 reports useful summaries to assess the models' performances in estimating N_{super}. The MAE and CIW are scale-dependent measures and do not allow for scenario comparisons. Therefore, we consider relative versions of these measures by dividing both by the corresponding expected value of N_{super}.

The RPT model returns the lowest relative MAEs in all the scenarios that involve more than 1 year of observation, whereas its competitors are associated with larger errors, which mostly increase with T. The WAIC seems to fail in selecting the RPT model when the number of sampling occasions is rather small (i.e., $T = 10$) by attaining the lowest values in correspondence of models that do not account for capture heterogeneity (i.e., M_1, M_5–M_7 in 52% of the replicas). The inability to identify the RPT structure in such a short-term scenario is not surprising, as it would be extremely complicated (if not impossible) to discern the effect of survival and capture probability when the occasions span only 1 year. However, as T increases, the WAIC tends to favor the true model, yielding the lowest median score and selecting it (i.e., returning the lowest WAIC) in most replicas. The classification performances of the RPT model are quite good in all the scenarios, with the mAUC improving as T increases. Notably, the resulting median mAUC is always ≥ 0.82 and above 0.95 when a year change occurs. Furthermore, by assigning the labels to each encountered individual according to the maximum a posteriori (MAP) rule, the median accuracy (across replicas) lies between 75% and 95% in all scenarios (once again improving as T increases).

Table 2 shows the RPT model performance in estimating some time-constant parameters. The estimates of the undetectability and the survival parameters have a very small MAE, although they fail to attain the nominal credible interval coverage of 95% in most scenarios. This indicates a good accuracy of the point estimates associated with over-confidence (visible in the low average CIWs), thus reducing the nominal coverage. Nonetheless, it settles to a fair and acceptable level. The result is not particularly surprising because these component-specific parameters govern latent ecological processes and are potentially mutually confounded.

Finally, we observe that the two survival probabilities (i.e., ϕ_T and ϕ_{NT}) are well separated in all scenarios, with an *OV* equal to 0.078 when $T = 10$ and equal to 0 for larger T.

5 | REAL DATA ANALYSIS

We apply the RPT model to estimate the total population size of the common bottlenose dolphins inhabiting the Tiber River estuary, as introduced in Section 2. The data are the detection histories of $D = 195$ well-marked dolphins that have been sighted in the area between June 2018 and November 2020, for a total of $T = 87$ occasions. After some preliminary runs with different values of M, we finally set $M - D = 500$ rows of pseudoindividuals (i.e., with null capture histories),

TABLE 1 Estimates of relative MAE (MAE_{rel}), coverage (Cov.), and relative average width of the 95% credible intervals (CIW_{rel}) for N_{super}, median WAIC, and percentage of times each competing model has achieved best WAIC (%waic). All these summaries have been obtained when data are simulated from the RPT model.

Model	T = 10					T = 20					T = 30					T = 40				
	MAE_{rel}	Cov.	CIW_{rel}	WAIC	%waic	MAE_{rel}	Cov.	CIW_{rel}	WAIC	%waic	MAE_{rel}	Cov.	CIW_{rel}	WAIC	%waic	MAE_{rel}	Cov.	CIW_{rel}	WAIC	%waic
RPT	0.07	0.96	0.34	1433.96	6	0.04	0.96	0.17	2953.01	50	0.04	0.86	0.15	4353.83	86	0.04	0.92	0.15	5840.58	88
$M_1 : \{\rho, \phi, p_t\}$	0.06	0.70	0.16	1458.18	2	0.09	0.28	0.11	3175.72	–	0.13	0	0.06	4875.18	–	0.14	0	0.07	6668.57	–
$M_2 : \{\rho_{t\times h_2}, \phi, p_{t+h_2}\}$	0.09	0.94	0.41	1448.27	–	0.04	0.92	0.19	2991.11	–	0.10	0.10	0.10	4495.79	–	0.10	0.24	0.13	6136.03	–
$M_3 : \{\rho_{t\times h_3}, \phi, p_{t+h_3}\}$	0.17	0.88	0.68	1435.05	4	0.04	0.94	0.27	3012.28	–	0.07	0.60	0.15	4532.30	–	0.09	0.42	0.15	6133.42	–
$M_4 : \{\rho_{t\times h_4}, \phi, p_{t+h_4}\}$	0.17	0.84	0.66	1439.58	8	0.07	0.92	0.28	2985.86	–	0.06	0.70	0.18	4510.22	–	0.07	0.64	0.21	6169.65	–
$M_5 : \{\rho_{t\times h_2}, \phi_{h_2}, p_t\}$	0.07	0.90	0.26	1448.64	8	0.08	0.90	0.27	3089.37	–	0.08	0.96	0.27	4710.57	–	0.10	0.76	0.26	6436.28	–
$M_6 : \{\rho_{t\times h_3}, \phi_{h_3}, p_t\}$	0.06	0.90	0.26	1438.10	10	0.09	0.90	0.26	3071.13	–	0.08	0.96	0.26	4706.34	–	0.10	0.74	0.27	6426.04	–
$M_7 : \{\rho_{t\times h_4}, \phi_{h_4}, p_t\}$	0.06	0.92	0.26	1435.73	32	0.09	0.84	0.27	3094.54	–	0.08	0.92	0.26	4725.57	–	0.11	0.72	0.27	6452.76	–
$M_8 : \{[\rho_{t\times h}, \phi_{t\times h}, p_{t\times h}]2\}$	0.16	0.90	0.67	1440.04	14	0.05	0.96	0.34	3041.38	10	0.11	0.68	0.48	4668.08	–	0.17	0.64	0.54	6453.97	–
$M_9 : \{[\rho_{t\times h}, \phi_{t\times h}, p_{t\times h}]3\}$	0.21	0.84	0.80	1430.92	4	0.30	0.62	0.76	2967.91	18	0.35	0.40	0.73	4444.86	6	0.41	0.20	0.64	6071.77	8
$M_{10} : \{[\rho_{t\times h}, \phi_{t\times h}, p_{t\times h}]4\}$	0.25	0.78	0.82	1437.25	12	0.33	0.52	0.81	2951.84	22	0.38	0.40	0.74	4424.14	8	0.41	0.14	0.59	6055.66	4

TABLE 2 Estimates of MAE, coverage (Cov.), and average width of the 95% credible intervals (CIW) for some time-constant parameters of model RPT. All these summaries have been obtained when data are simulated from the RPT model.

	ϕ_T			ϕ_{NT}			δ			μ		
T	MAE	Cov.	CIW	MAE	Cov.	CIW	MAE	Cov.	CIW	MAE	Cov.	CIW
10	0.32	0.82	0.86	0.05	0.62	0.09	0.05	0.94	0.28	0.25	0.94	0.97
20	0.10	0.88	0.35	0.01	0.68	0.01	0.03	0.82	0.11	0.13	0.82	0.44
30	0.08	0.80	0.25	0.002	0.76	0.007	0.03	0.62	0.08	0.08	0.90	0.32
40	0.07	0.82	0.20	0.002	0.80	0.01	0.02	0.74	0.06	0.06	0.96	0.27

TABLE 3 WAIC, estimated total superpopulation abundance and by group (95% credible intervals).

Model	WAIC	$\hat{N}_{super}$
RPT	4725.1	311 (266, 373)
$\{[\rho_{t \times h}, \phi_{t \times h}, p_{t \times h}]_3\}$	4801.7	600 (467, 669)
$\{[\rho_{t \times h}, \phi_{t \times h}, p_{t \times h}]_4\}$	4811.9	594 (409, 664)
$\{\rho_{t \times h_2}, \phi_{h_2}, p_t\}$	4844.6	341 (287, 418)
$\{\rho_{t \times h_3}, \phi, p_{t+h_3}\}$	4870.3	274 (243, 299)
$\{\rho_{t \times h_2}, \phi, p_{t+h_2}\}$	4877.9	276 (240, 307)
$\{\rho_{t \times h_4}, \phi_{h_4}, p_t\}$	4893.1	352 (281, 458)
$\{\rho_{t \times h_4}, \phi, p_{t+h_4}\}$	4903.8	271 (244, 306)
$\{\rho_{t \times h_3}, \phi_{h_3}, p_t\}$	4932.2	334 (269, 441)
$\{[\rho_{t \times h}, \phi_{t \times h}, p_{t \times h}]_2\}$	5088.5	356 (267, 628)
$\{\rho_t, \phi, p_t\}$	5108.7	230 (216, 242)

thus yielding $M = 695$. Other choices of M led to similar results, with larger values of M only straining the computational burden, in terms of runtime and storage.

We run two parallel chains, each with 20,000 iterations with a burn-in of 5000 iterations and no thinning. We compare the performances of the RPT model with the wide range of alternative models illustrated in Section 4, using the same prior setting specified in the simulation study for all the considered models. Table 3 reports the results on abundance estimation along with the WAIC associated with each competing model. We notice that the RPT model does yield the lowest WAIC score. Interestingly, the second best choice according to the WAIC is model $\{[\rho_{t \times h}, \phi_{t \times h}, p_{t \times h}]_3\}$, which indeed resembles the same structure as the RPT model but without allowing for mixture components sharing common parameters. This lack of parsimony results in an overestimation of the individuals in the superpopulation, similar to what has been observed in the simulation study (cf. Section 4). Thus, these results suggest the presence of unobserved heterogeneity in the population, which seems to be better described by the more parsimonious structure of the RPT model than by a generic three-group specification.

The two annual survival probabilities are well separated with an $OV = 0$ and posterior estimates are $\hat{\phi}_T = 1.06 \times 10^{-8}$ ($CI_{0.95} = [0, 2.8 \times 10^{-6}]$) for the group of transient individuals and $\hat{\phi}_{NT} = 0.71$ ($CI_{0.95} = [0.62, 0.80]$) for the resident and part-time individuals. Notice that the estimated survival parameter $\hat{\phi}_T$ is of little interpretability on the annual scale. However, it corresponds to a probability of 0.73 on the weekly scale and 0.26 on the monthly scale. The average capture probability of the resident and transient individuals is $\hat{p}_{NP} = \text{logit}^{-1}(\hat{\mu}) \approx 0.19$; the corresponding temporal variations, captured by τ_t, result in the time-dependent posterior distributions of $p_{NP,t}$ reported in Figure F1 in the Appendix.

The parameter δ regulating the undetectability of the part-time individuals is estimated to be $\hat{\delta} \approx 0.74$. This means that individuals in that group are present in the area approximately for the 26% of their lifetime. Although the estimates of the recruitment parameters have little interpretation in the Royle and Dorazio (2008)'s considered framework, it is worth noticing that, on average, the recruitment probabilities are higher during the first year for the resident individuals, while approximately constant for part-time and transient individuals (cf. Figure F2 in the Appendix). This is a model artifact motivated by all the individuals already present in the population before the start of the survey (i.e., mostly residents) and that are *virtually* recruited on the first occasion. If we consider the second year only (most reliable in terms of recruitment probability estimation and individual classification), the average recruitment probabilities are approximately
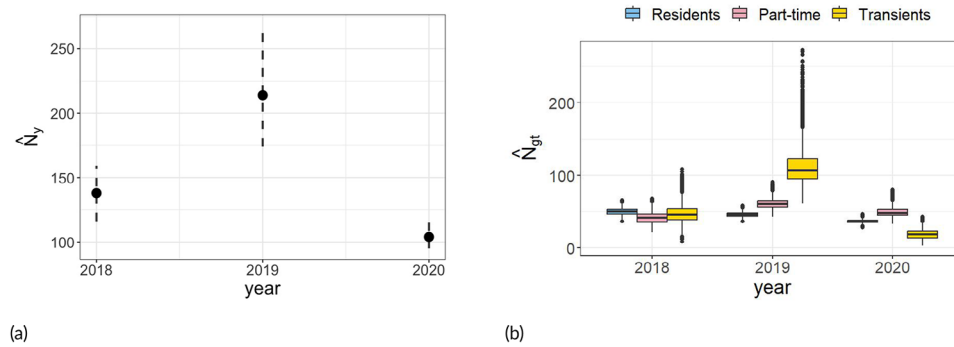
FIGURE 3 Point estimates and posterior 95% credible intervals of the yearly superpopulation size (a) and by group (b).

7×10^{-4}, 1×10^{-3}, and 6×10^{-3} for the three groups, respectively. This aligns with our expectations, as recruiting more stable individuals is a slower process than recruiting less stable ones.

The final abundance estimate of the superpopulation in the whole observation window is of $\hat{N}_{super} = 311$ ($CI_{0.95} = [266, 373]$), with yearly variations $\hat{N}_y$ ($y = 2018, 2019, 2020$) that show a peak in 2019 and a decrease in the last year of observation (see Figure 3a). It is, however, interesting to look at Figure 3(b), where we report the yearly abundance estimates by group to better understand the behavior of the aggregated yearly pattern in Figure 3a).

Indeed, it seems that transients' abundance ($\hat{N}_T = 174$ ($CI_{0.95} = [123, 236]$)) is the main factor affecting the overall yearly counts, especially in 2019 and 2020. They show a substantial decrease in 2020, whereas the abundances of residents ($\hat{N}_R = 57$ ($CI_{0.95} = [48, 65]$)) and part-time ($\hat{N}_P = 81$ ($CI_{0.95} = [55, 101]$)) group tend to remain stable across years. Recall that these estimates only refer to well-marked individuals, who represent only a proportion of the whole population ($\approx 60\%$ of all sighted individuals across the 3 years).[3] The posterior distributions of the yearly sizes of the three groups were obtained by counting all the individuals present in the population at each iteration by year and group.

By assigning the 195 well-marked individuals observed between June 2018 and November 2020 to a single group according to the *MAP* allocation, we have that 51 are assigned to the group of residents, 54 to the group of part-time, and 90 to the group of transients. Notice that the MAP is a straightforward and well-established method to attain classification in finite-mixture modeling. However, it is known to be suboptimal in some contexts (McLachlan et al., 2019; Stephens, 2000) and other appealing methods have been proposed in the recent literature (e.g., Wade & Ghahramani, 2018). The problem of summarizing membership probabilities into a crisp classification is a challenging and long-debated issue on which there is no general agreement. Nevertheless, results from the simulation study (see Section 4) showed that the MAP is sufficiently reliable in the proposed framework. On the other hand, we are aware that any procedure deriving a crisp classification from membership probabilities oversimplifies the complexity of the inferred results as it does not give any information about the strength each individual is assigned to a specific group. The latter is one of the main advantages of the soft clustering returned by FMMs, and we wish to exploit its fuzziness to quantify how decisively each individual is assigned to one group or the other. For instance, 90% of the individuals classified into the Resident group have been assigned to it with a probability greater than 0.9. This probability is greater than 0.54 for the Part-Times and 0.53 for the Transient.

In this regard, it is possible to visualize the classification results in Figure 4 using a *ternary diagram*, which allows quantifying the probability that each individual belongs to each group. The figure highlights three well-distinguishable capture history patterns that again comply with the typical RPT behavior. Residents are available in the area for the whole study period and are spotted very frequently in subsequent sampling occasions; their capture histories are very informative. Hence, their identification is clear-cut (intense color) in most cases. Part-time individuals are available in the area for most of the study period and are frequently spotted, but not as often as residents. Their classification is crisp if they have been encountered for the first time toward the beginning of the study period, while it is dicey when they have been encountered toward the end for the first time. Finally, the transients show short capture histories with few captures that never cross 2 years; this reflects that they are spotted only a few times and do not visit (*survive*) the area for long. Given the little information provided by their capture histories, their classification is slightly vaguer compared to the other groups, especially for individuals only observed in the last sampling occasions.

Finally, it is important to remark that the proportions of the three groups do not align with the estimated weights $\hat{w} = (0.09, 0.39, 0.52)$. This is because the prior weights are not a good indicator of the true impact of each component

---

[3] Inferring the overall size from the well-marked abundance estimate is not the focus of this paper. For details on this passage, see Wilson et al. (1999).
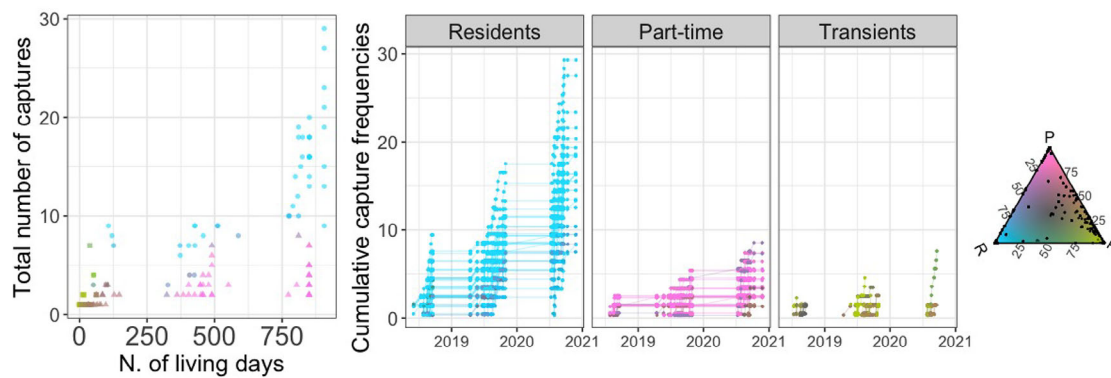
FIGURE 4 Individual cumulative frequencies of capture for all the encountered individuals divided into the three groups defined by the RPT model. Posterior allocation was based on MAP.

in the superpopulation, but only of their impact on the augmented dataset (as mentioned in Section 4). However, we can infer the composition of the $N - D$ uncaught individuals, namely, those individuals that were recruited (i.e., not pseudoindividuals) but were never captured. The model allocates (on average) the uncaught individuals to the resident group for the 2%, the part-time group for the 15%, and the transient group for the 83%. This result complies with the short-term survival of the transient individuals and the more elusive nature of the part-time ones. On the other hand, the residents are always present in the area, and therefore, more susceptible to capture in a broad sense. The results suggest that most of the resident individuals (98%) have already been observed and documented, a result that is in line with the progressive decrease of the discovery rate highlighted in Section 2 (cf. Figure 1a).

6 | DISCUSSION

Estimating the abundance of marine wildlife species is a challenging but critical activity that can tell much about the undergoing ecological processes. Thus, combining high-quality data with solid analytical approaches is essential to improve our knowledge of these dynamics and increase the potential for management actions (Lin et al., 2022; Vella et al., 2021). In this work, we pursued a Bayesian estimation of the size of a common bottlenose dolphin population, which is organized into three groups with different residency and site-fidelity patterns. Accounting for such unobserved heterogeneity is a very common problem in the environmental literature. However, few papers approach the problem from the Bayesian perspective and develop ad-hoc solutions based on prior scientific knowledge of the population of interest.

We proposed a parsimonious specification of an FMM within the PX-DA setting for CR analysis, which we named RPT. This specification reflects the typical bottlenose dolphin residency pattern, with individuals showing high, partial, or low site fidelity (Dinis et al., 2016; Haughey et al., 2020; Hunt et al., 2017; La Manna et al., 2022). Its adoption, characterized by fewer free parameters, simplifies the identification of all the model components compared to more generic and flexible alternatives. Furthermore, while the application of FMMs to bring evidence about the true existence of different population groups has sometimes been discouraged (Pledger, 2000; Pledger et al., 2003), here we have shown how they can be exploited to identify classes of individuals sharing similar profiles whenever strong scientific evidence of population groups' existence is available.

We devoted part of the model description to the discussion of a suitable prior elicitation preventing the label-switching issue of FMM. Formal derivations have enriched the results on constrained Beta priors from Alaimo Di Loro et al. (2022). The performances of our proposal have been assessed through a simulation study that considered scenarios with the increasing length of the observation window and, hence, of the capture histories (i.e., mimicking continuous monitoring of a population of interest over time). When data were generated from the RPT model, we evaluated the influence of the number of sampling occasions on the estimates' quality. As expected, yielded better accuracy and coverage thanks to the larger sample size. Comparison with alternative model versions over many replicas showed that the model estimation exhibited satisfying and robust performances when the observation window was long enough (more than 1 year of monitoring).

We estimated the RPT model and other alternatives on the data that motivated our work. The results showed that, in terms of the WAIC score, our model outplays several well-established competitors in the class of JS-type models. The results correspond to biologically meaningful findings and align with previous research work. The estimated abundance is of $\hat{N}^*_{super} = 311$, of which 51 are estimated to be resident members of the population. This quantity is particularly important as it is a proxy of the breeding population in the area. The yearly trend shows an increase between 2018 and 2019 and a decrease between 2019 and 2020. While the estimated overall abundances are close to the ones obtained by Pace et al. (2021), the trend is slightly in contrast with their results (nonsignificant increasing trend between 2019 and 2020). This difference must be due to the better quantification of the overall uncertainty in our analysis. This is particularly important in this context as the major interest lies in estimating a reasonable range that accounts for both the *worst* and *best* case scenario. Indeed, we estimate that the driving factor of the variations in the population size across the years is due to the abundance variation of the transient individuals, the proportion of which is dynamic and not static. Changes in the presence of such dolphins in an area may reflect changes in driving features such as habitat quality, prey distribution, or anthropogenic disturbance. The year 2020 was the COVID-19 pandemic, with changes in ecological conditions and human pressures on coastal marine waters (Carome et al., 2022). Although difficult to assess, the influence of these factors on the number of transient individuals cannot be ruled out, as several dolphins may have been induced to reduce their mobility toward the study area by improved conditions in their native habitat.

One basic assumption of CR experiments that we do not drop is that captures of different individuals are independent. However, it is well known that bottlenose dolphin populations can form structured societies with complex social networks (Pace et al., 2022a). In future studies, we would like to drop the independence assumption and include information on the population's social structure and the consequent statistical dependence in the capture histories. Including the effect of external or individual covariates would also be interesting. For instance, it is possible to recognize the gender and age class in high-quality pictures. This partial information may be incorporated in the Bayesian framework and could help for a better assessment, for example, of the membership of different individuals in different groups or of their marking probability (Wu et al., 2021). As kindly suggested by a reviewer, the CR model to estimate the abundance of the bottlenose dolphins could be extended by adopting a stopover model (Pledger et al., 2009; Worthington et al., 2019), which allows capture and survival probabilities to depend both on time and time since arrival in the population. In that case, the model could be formulated as a multistate HMM (Worthington et al., 2019), where the different states refer to the different groups of individuals in the population. Last but not least, it would be extremely interesting to conduct the survey on a larger spatial scale and include external information to account for the spatial heterogeneity (Wu & Holan, 2017).

CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available in the supplementary material of this article.

OPEN RESEARCH BADGES
This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge "**Reproducible Research**" for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

ORCID
Gianmarco Caruso https://orcid.org/0000-0001-8765-0003
Pierfrancesco Alaimo Di Loro https://orcid.org/0000-0002-6075-3659

Marco Mingione https://orcid.org/0000-0002-5662-3499
Luca Tardella https://orcid.org/0000-0002-5703-3909
Daniela Silvia Pace https://orcid.org/0000-0001-5121-7080
Giovanna Jona Lasinio https://orcid.org/0000-0001-8912-5018

REFERENCES

Alaimo Di Loro, P., Caruso, G., Mingione, M., Jona Lasinio, G., & Tardella, L. (2022). Specification of informative priors for capture-recapture finite mixture models. In *Book of Short Paper – SIS 2022*. Pearson.

Altieri, L., Farcomeni, A., & Alunni Fegatelli, D. A. (2023). Continuous time-interaction processes for population size estimation, with an application to drug dealing in Italy. *Biometrics*, *79*, 1254.

Alunni Fegatelli, D., & Tardella, L. (2013). Improved inference on capture recapture models with behavioural effects. *Statistical Methods & Applications*, *22*, 45–66.

Amstrup, S. C., McDonald, T. L., & Manly, B. F. (2010). *Handbook of capture-recapture analysis*. Princeton University Press.

Arnold, R., Hayakawa, Y., & Yip, P. (2010). Capture–recapture estimation using finite mixtures of arbitrary dimension. *Biometrics*, *66*, 644–655.

Böhning, D., Bunge, J., & van der Heijden, P. G. M. (2018). *Capture-recapture methods for the social and medical sciences*. CRC Press Boca Raton.

Böhning, D., & van der Heijden, P. G. M. (2009). Recent developments in life and social science applications of capture-recapture methods. *AStA Advances in Statistical Analysis*, *93*, 1.

Böhning, D., Rocchetti, I., Maruotti, A., & Holling, H. (2020). Estimating the undetected infections in the covid-19 outbreak by harnessing capture–recapture methods. *International Journal of Infectious Diseases*, *97*, 197–201. URL: https://www.sciencedirect.com/science/article/pii/S1201971220304446

Bravington, M. V., Skaug, H. J., & Anderson, E. C. (2016). Close-kin mark-recapture. *Statistical Science*, *31*, 259–274.

Brittain, S., & Böhning, D. (2009). Estimators in capture–recapture studies with two sources. *AStA Advances in Statistical Analysis*, *93*, 23–47.

Brooks, S., King, R., & Morgan, B. (2004). A Bayesian approach to combining animal abundance and demographic data. *Animal Biodiversity and Conservation*, *27*, 515–529.

Brooks, S. P., Catchpole, E. A., Morgan, B. J. T., & Barry, S. C. (2000). On the Bayesian analysis of ring-recovery data. *Biometrics*, *56*, 951–956.

Carome, W., Slooten, E., Rayment, W., Webster, T., Wickman, L., Brough, T., & Dawson, S. M. (2022). A long-term shift in the summer distribution of Hector's dolphins is correlated with an increase in cruise ship tourism. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *32*, 1660–1674.

Chao, A., Tsay, P. K., Lin, S. H., Shau, W. Y., & Chao, D. Y. (2001). The applications of capture-recapture models to epidemiological data. *Statistics in Medicine*, *20*, 3123–3157.

Chung, H., Loken, E., & Schafer, J. L. (2004). Difficulties in drawing inferences with finite-mixture models: A simple example with a simple solution. *The American Statistician*, *58*, 152–158.

Clark, J. S., & Gelfand, A. E. (2006). *Hierarchical modelling for the environmental sciences: Statistical methods and applications*. OUP Oxford.

Di Cecco, D., Di Zio, M., & Liseo, B. (2020). Bayesian latent class models for capture–recapture in the presence of missing data. *Biometrical Journal*, *62*, 957–969.

Diebolt, J., & Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society: Series B (Methodological)*, *56*, 363–375.

Dinis, A., Alves, F., Nicolau, C., Ribeiro, C., Kaufmann, M., Ca nadas, A., & Freitas, L. (2016). Bottlenose dolphin *Tursiops truncatus* group dynamics, site fidelity, residency and movement patterns in the Madeira Archipelago (North-East Atlantic). *African Journal of Marine Science*, *38*, 151–160.

Dorazio, R. M. (2020). Objective prior distributions for Jolly-Seber models of zero-augmented data. *Biometrics*, *76*, 1285–1296.

Dorazio, R. M., & Royle, J. A. (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics*, *59*, 351–364.

Farcomeni, A. (2022). How many refugees and migrants died trying to reach europe? Joint population size and total estimation. *The Annals of Applied Statistics*, *16*, 2339–2351.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*, 997–1016.

Ghosh, S. K., & Norris, J. L. (2005). Bayesian capture-recapture analysis and model selection allowing for heterogeneity and behavioral effects. *Journal of Agricultural, Biological, and Environmental Statistics*, *10*, 35–49.

Gimenez, O., Cam, E., & Gaillard, J. M. (2018). Individual heterogeneity and capture–recapture models: What, why and how? *Oikos*, *127*, 664–686.

Guéry, L., Descamps, S., Pradel, R., Hanssen, S. A., Erikstad, K. E., Gabrielsen, G. W., Gilchrist, H. G., & Bêty, J. (2017). Hidden survival heterogeneity of three Common eider populations in response to climate fluctuations. *Journal of Animal Ecology*, *86*, 683–693.

Hand, D. J., & Till, R. J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning*, *45*, 171–186.

Haughey, R., Hunt, T., Hanf, D., Rankin, R. W., & Parra, G. J. (2020). Photographic capture-recapture analysis reveals a large population of Indo-Pacific bottlenose dolphins (Tursiops aduncus) with low site fidelity off the North West Cape, Western Australia. *Frontiers in Marine Science*, *6*, 781.

Hunt, T. N., Bejder, L., Allen, S. J., Rankin, R. W., Hanf, D., & Parra, G. J. (2017). Demographic characteristics of Australian humpback dolphins reveal important habitat toward the southwestern limit of their range. *Endangered Species Research*, *32*, 71–88.

Kendall, W. L., Nichols, J. D., & Hines, J. E. (1997). Estimating temporary emigration using capture–recapture data with pollock's robust design. *Ecology*, *78*, 563–578.

King, R., & McCrea, R. (2019). Chapter 2 - Capture–recapture methods and models: Estimating population size. In A. S. Srinivasa Rao & C. Rao (Eds.), *Integrated population biology and modeling, Part B*, Vol. 40 of Handbook of Statistics (pp. 33–83). Elsevier.

La Manna, G., Rako-Gospić, N., Pace, D. S., Bonizzoni, S., Di Iorio, L., Polimeno, L., Perretti, F., Ronchetti, F., Giacomini, G., Pavan, G., Pedrazzi, G., Labach, H., & Ceccherelli, G. (2022). Determinants of variability in signature whistles of the Mediterranean common bottlenose dolphin. *Scientific Reports*, *12*, 1–16.

Le Cren, E. D. (1965). A note on the history of mark-recapture population estimates. *Journal of Animal Ecology*, *34*, 453–454.

Lin, W., Zheng, R., Liu, B., Chen, S., Lin, M., Liu, M., Liu, W., & Li, S. (2022). Low survivals and rapid demographic decline of a threatened estuarine Delphinid. *Frontiers in Marine Science*, *9*.

Maruotti, A., Böhning, D., Rocchetti, I., & Ciccozzi, M. (2023). Estimating the undetected infections in the monkeypox outbreak. *Journal of Medical Virology*, *95*, e28099.

Matechou, E., & Argiento, R. (2023). Capture-recapture models with heterogeneous temporary emigration. *Journal of the American Statistical Association*, *118*, 56.

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*, 120–135.

McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite mixture models. *Annual Review of Statistics and its Application*, *6*, 355–378.

Mena, R. H., & Walker, S. G. (2015). On the Bayesian mixture model and identifiability. *Journal of Computational and Graphical Statistics*, *24*, 1155–1169.

Morin, D. J., Kelly, M. J., & Waits, L. P. (2016). Monitoring coyote population dynamics with fecal DNA and spatial capture–recapture. *The Journal of Wildlife Management*, *80*, 824–836.

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

Pace, D. S., Di Marco, C., Giacomini, G., Ferri, S., Silvestri, M., Papale, E., Casoli, E., Ventura, D., Mingione, M., Alaimo Di Loro, P., Jona Lasinio, G., & Ardizzone, G. (2021). Capitoline dolphins: Residency patterns and abundance estimate of *Tursiops truncatus* at the Tiber River Estuary (Mediterranean Sea). *Biology*, *10*, 275.

Pace, D. S., Ferri, S., Giacomini, G., Di Marco, C., Papale, E., Silvestri, M., Pedrazzi, G., Ventura, D., Casoli, E., & Ardizzone, G. (2022a). Resources and population traits modulate the association patterns in the common bottlenose dolphin living nearby the Tiber river estuary (Mediterranean Sea). *Frontiers in Marine Science*, *9*, 1506.

Pace, D. S., Panunzi, G., Arcangeli, A., Moro, S., Jona-Lasinio, G., & Martino, S. (2022b). Seasonal distribution of an opportunistic apex predator (Tursiops truncatus) in marine coastal habitats of the Western Mediterranean Sea. *Frontiers in Marine Science*, *9*, 1785.

Pace, D. S., Tumino, C., Silvestri, M., Giacomini, G., Pedrazzi, G., Pavan, G., Papale, E., Ceraulo, M., Buscaino, G., & Ardizzone, G. (2022c). Bray-call sequences in the Mediterranean Common Bottlenose Dolphin *Tursiops truncatus* acoustic repertoire. *Biology*, *11*, 367.

Papale, E., Alonge, G., Caruso, F., Grammauta, R., Mazzola, S., Mussi, B., Pace, D. S., & Buscaino, G. (2021). The higher, the closer, the better? Influence of sampling frequency and distance on the acoustic properties of short-beaked common dolphins burst pulses in the Mediterranean Sea. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *31*, 51–60.

Pastore, M. (2018). Overlapping: A R package for estimating overlapping in empirical distributions. *Journal of Open Source Software*, *3*, 1023.

Pastore, M., & Calcagnì, A. (2019). Measuring distribution similarities between samples: A distribution-free overlapping index. *Frontiers in Psychology*, *10*, 1089.

Pledger, S. (2000). Unified maximum likelihood estimates for closed capture–recapture models using mixtures. *Biometrics*, *56*, 434–442.

Pledger, S. (2005). The performance of mixture models in heterogeneous closed population capture–recapture. *Biometrics*, *61*, 868–873.

Pledger, S., Efford, M., Pollock, K., Collazo, J., & Lyons, J. (2009). Stopover duration analysis with departure probability dependent on unknown time since arrival. In D. L. Thomson, E. G. Cooch, M. J. Conroy, (Eds.), *Modeling Demographic Processes in Marked Populations*, Vol. 3 Environmental and Ecological Statistics, Springer, Boston, MA.

Pledger, S., Pollock, K. H., & Norris, J. L. (2003). Open capture-recapture models with heterogeneity: I. Cormack-Jolly-Seber model. *Biometrics*, *59*, 786–794.

Pledger, S., Pollock, K. H., & Norris, J. L. (2010). Open capture–recapture models with heterogeneity: II. Jolly–Seber model. *Biometrics*, *66*, 883–890.

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

Richardson, S., & Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *59*, 731–792.

Royle, J. A., & Dorazio, R. M. (2008). *Hierarchical modeling and inference in ecology: The analysis of data from populations, metapopulations and communities*. Elsevier.

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

Royle, J. A., Karanth, K. U., Gopalaswamy, A. M., & Kumar, N. S. (2009). Bayesian inference in camera trapping studies for a class of spatial capture–recapture models. *Ecology*, *90*, 3233–3244.

Schwarz, C. J., & Arnason, A. N. (1996). A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics*, *52*, 860–873.

Silverman, B. W. (2020). Multiple-systems analysis for the quantification of modern slavery: Classical and Bayesian approaches. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *183*, 691–736.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*, 795–809.

Turek, D., Wehrhahn, C., & Gimenez, O. (2021). Bayesian non-parametric detection heterogeneity in ecological models. *Environmental and Ecological Statistics*, *28*, 355–381.

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

Vella, A., Murphy, S., Giménez, J., de Stephanis, R., Mussi, B., Vella, J. G., Larbi Doukara, K., & Pace, D. S. (2021). The conservation of the endangered Mediterranean common dolphin (Delphinus delphis): Current knowledge and research priorities. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *31*, 110–136.

Wade, S., & Ghahramani, Z. (2018). Bayesian cluster analysis: Point estimation and credible balls (with discussion). *Bayesian Analysis*, *13*, 559–626.

Watanabe, S., & Opper, M. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571.

Wilson, B., Hammond, P. S., & Thompson, P. M. (1999). Estimating size and assessing trends in a coastal bottlenose dolphin population. *Ecological Applications*, *9*, 288–300.

Worthington, H., McCrea, R., King, R., & Griffiths, R. (2019). Estimating abundance from multiple sampling capture-recapture data via a multi-state multi-period stopover model. *Annals of Applied Statistics*, *13*, 2043.

Wu, G., & Holan, S. H. (2017). Bayesian hierarchical multi-population multistate Jolly–Seber models with covariates: Application to the Pallid Sturgeon Population Assessment Program. *Journal of the American Statistical Association*, *112*, 471–483.

Wu, G., Holan, S. H., Avril, A., & Waldenström, J. (2021). A Bayesian semiparametric Jolly–Seber model with individual heterogeneity: An application to migratory mallards at stopover. *The Annals of Applied Statistics*, *15*, 813–830.

Zhou, M., McCrea, R. S., Matechou, E., Cole, D. J., & Griffiths, R. A. (2019). Removal models accounting for temporary emigration. *Biometrics*, *75*, 24–35.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Caruso, G., Alaimo Di Loro, P., Mingione, M., Tardella, L., Pace, D. S., & Jona Lasinio, G. (2024). Finite mixtures in capture–recapture surveys for modeling residency patterns in marine wildlife populations. *Biometrical Journal*, *66*, 2200350. https://doi.org/10.1002/bimj.202200350

APPENDIX A: MULTISTATE INTERPRETATION OF THE ECOLOGICAL PROCESS

The ecological model described by Royle and Dorazio (2008) can be also seen as a multistate process. At each occasion t, each individual in the augmented matrix can be in one and only one of the following three states:

1. It has never been part of the population.
2. It is part of the population.
3. It was part of the population, but now it is not.

We first notice that when an individual becomes part of the population, it cannot be recruited any more: following the notation introduced in Section 3.1, this implies that for $t > 1$, r_{it} and z_{it} cannot be simultaneously equal to 1.

In the JS modeling framework, individuals who leave the population cannot return to it. Hence, state 3 is an absorbing state. Let us momentarily ignore the population heterogeneity (clustering structure) for the sake of clarity. If we allow temporal heterogeneity, then the transition probability matrix associated with the three states at times $t = 2, \dots, T$ is:

$$
\begin{array}{c|ccc}
 & 1 & 2 & 3 \\
\hline
1 & 1 - \rho_t & \rho_t & 0 \\
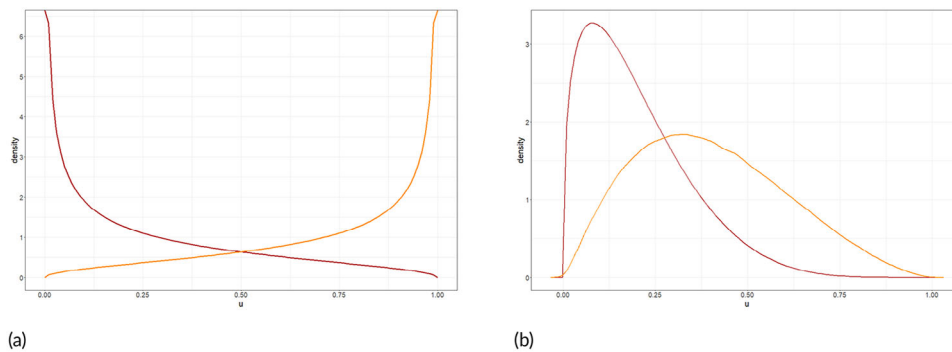2 & 0 & \phi_t & 1 - \phi_t \\
3 & 0 & 0 & 1
\end{array}
$$

(a)   (b)

FIGURE B1 (a) Marginal priors induced on u_1 (solid red) and u_2 (dashed orange) by the conditional specification $u_1 \sim Beta(.5, 1.5)$, $u_2|u_1 \sim tBeta(1, .5, u_1, 1)$. (b) Marginal priors induced on u_1 (solid red) and u_2 (dashed orange) with $\mathbb{E}[u_1] = 0.2$, $\mathbb{E}[u_2] = 0.4$ with $\mathbb{V}[u_1] = 0.02$ and $\mathbb{V}[u_2] = 0.04$ by the conditional prior specification $u_1 \sim Beta(1.4, 5.6)$, $u_2|u_1 \sim rBeta(0.826, 2.478, u_1, 1)$.

where rows and columns represent the states at time t and $t + 1$, respectively. At period $t = 1$, all individuals can be recruited in the population. As t increases, more and more individuals enter the population or, equivalently, are *removed* from state 1. All the observed individuals will eventually be recruited in the population before time T, but not all of them will leave it (may have survived to future, unobserved, periods). Remember that individuals are exposed to capture, with probability p_t, only during their transitory stay in state 2. Hence, also a portion of never observed individuals may have been recruited into the population at some point without ever being captured. They represent the unknown part we aim to estimate.

APPENDIX B: MORE DETAILS ON THE PRIOR SPECIFICATION

The standard solution is that of concatenating conditionally specified Uniform distributions as follows:

$$u_1 \sim Unif(0, 1), \quad u_j|u_{j-1} \sim Unif(u_{j-1}, 1) \qquad j = 2, \dots, G, \tag{B1}$$

where the u_j's are, for instance, the survival probabilities (Turek et al., 2021). While effective in imposing the constraint, the priors in (B1) do not allow for the inclusion of previous information that can ease parameters' identification. Alaimo Di Loro et al. (2022) explore alternative conditional prior specification that, while implementing the ordering constraint, allows to control for the shape and first moments of the induced marginal prior distributions:

$$\pi_{u_g}(u_g) = \int_{S_{g-1}} \pi_{u_1}(u_1) \prod_{j=2}^{g} \pi_{u_j|u_{j-1}}(u_j \mid u_{j-1}) \, du_1, \dots, u_{g-1},$$

where S_{g-1} is a simplex of order $g - 1$. Possible choices are the Beta and Truncated Beta distributions. The latter corresponds to the following set of prior distributions:

$$u_1 \sim Beta(\alpha_1, \beta_1), \qquad u_g|u_{g-1} \sim tBeta(\alpha_g, \beta_g; u_{g-1}, 1) \qquad g = 2, \dots, G, \tag{B2}$$

where $tBeta(\alpha_g, \beta_g; l, 1)$ denotes the Truncated Beta distribution in $(l, 1)$. Note that for $\alpha_g = \beta_g = 1$, we obtain (B1). When $G = 2$ and $\alpha_2 = 1$, the prior specification in (B2) induces the following marginal prior density on u_2:

$$\pi_{u_2}(u_2) = \frac{B(\alpha_1, \beta_1 - \beta_2)}{B(\alpha_1, \beta_1)} \beta_2 (1 - u_2)^{\beta_2 - 1} F_{Beta(\alpha_1, \beta_1 - \beta_2)}(u_2), \quad \text{with } \beta_1 > \beta_2, \tag{B3}$$

where $F_{Beta(\alpha_1, \beta_1 - \beta_2)}(\cdot)$ is the cdf of a $Beta(\alpha_1, \beta_1 - \beta_2)$. See Section B.1 for the formal proof. When $\alpha_1 = \beta_2 = k$ and $\beta_1 = k + 1$, we have that $\pi_{u_1}(u_1) = Beta(u_1 \mid k, k + 1)$ and the density in (B3) is a $Beta(k + 1, k)$. The two distributions are mirrored with respect to the vertical line $v = 0.5$ (see Figure B1a of the Appendix as an example). Hence, we have $\mathbb{E}[u_2] = 1 - \mathbb{E}[u_1]$, for all $k > 0$; if $k > 1$, we also have $Mode(u_2) = 1 - Mode(u_1)$. The low-parametrized structure induces well-

separated prior means or prior modes for u_1, u_2 marginal distributions, favoring the mixture components' separation. Alternative settings inducing well-separated modes in beta-type priors are reported in Section B.2 of the Appendix.

B.1 | The beta and truncated beta distribution

The density of a Beta random variable with shape α and rate β is:

$$Beta(z \mid \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot z^{\alpha-1} \cdot (1-z)^{\beta-1}, \quad z \in (0,1),$$

where $\Gamma(\cdot)$ is the Euler gamma function. The truncation requires normalizing the same distribution over the truncated domain. We say that $u_2|u_1$ is a truncated beta in $(u_1, 1)$ when:

$$\pi(u_2|u_1) = tBeta(\alpha_2, \beta_2, u_1, 1) = \frac{1}{B(\alpha_2, \beta_2)} \frac{u_2^{\alpha_2-1}(1-u_2)^{\beta_2-1}}{1 - F_{Beta(\alpha_2,\beta_2)}(u_1)} \mathbb{1}_{\{u_1, 1\}}(u_2),$$

with $F_{Beta(\alpha_2,\beta_2)}(\cdot)$ being the cdf of a $Beta(\alpha_2, \beta_2)$. Now, suppose that $u_1 \sim Beta(\alpha_1, \beta_1)$ and $u_2|u_1 \sim tBeta(\alpha_2, \beta_2, u_1, 1)$, then if $\alpha_2 = 1$, the marginal prior distribution induced on u_2 is given by:

$$
\begin{aligned}
\pi(u_2) &= \int_0^1 \pi(u_2|u_1)\pi(u_1)\,du_1 \\
&= \int_0^1 \frac{\beta_2(1-u_2)^{\beta_2-1}}{1 - F_{Beta(1,\beta_2)}(u_2)} \mathbb{1}_{\{u_1, 1\}}(u_2) \frac{1}{B(\alpha_1, \beta_1)} \times u_1^{\alpha_1-1}(1-u_1)^{\beta_1-1}\,du_1 \\
&= \frac{\beta_2(1-u_2)^{\beta_2-1}}{B(\alpha_1, \beta_1)} \int_0^{u_2} \frac{u_1^{\alpha_1-1}(1-u_1)^{\beta_1-1}}{1 - F_{Beta(1,\beta_2)}(u_1)}\,du_1 \\
&= \frac{\beta_2(1-u_2)^{\beta_2-1}}{B(\alpha_1, \beta_1)} \int_0^{u_2} u_1^{\alpha_1-1}(1-u_1)^{(\beta_1-\beta_2)-1}\,du_1 \\
&= \frac{B(\alpha_1, \beta_1-\beta_2)}{B(\alpha_1, \beta_1)} \beta_2(1-u_2)^{\beta_2-1} \int_0^{u_2} \frac{u_1^{\alpha_1-1}(1-u_1)^{(\beta_1-\beta_2)-1}}{B(\alpha_1, \beta_1-\beta_2)}\,du_1 \\
&= \frac{B(\alpha_1, \beta_1-\beta_2)}{B(\alpha_1, \beta_1)} \beta_2(1-u_2)^{\beta_2-1} F_{Beta(\alpha_1, \beta_1-\beta_2)}(u_2), \quad with\ \beta_1 > \beta_2,
\end{aligned}
$$

where we exploited the fact that $F_{Beta(1,b)}(z) = \int_0^z (1-t)^{b-1}\,dt = 1 - (1-z)^b$. Observe that the constraint $\beta_1 > \beta_2$ is essential to avoid the divergence of the beta function.

In particular, it is straightforward to show that if $\alpha_1 = \beta_2 = k$ and $\beta_1 = k+1$, that is, $u_1 \sim Beta(k, k+1)$ and $u_2|u_1 \sim tBeta(1, k, u_1, 1)$, then $u_2 \sim Beta(k+1, k)$. Notice that, in this particular case, the marginal distribution induced on u_2 is symmetrical with respect to the distribution of u_1 around the vertical line $v = 0.5$; equivalently, $u_2 \overset{d}{=} 1 - u_1$.

B.2 | The beta and restricted beta

An alternative conditional specification that allows for a properly informed marginal prior can be obtained using restricted beta distribution. Also known as *4-parameters Beta*, it is a Beta r.v. that has been shifted and scaled to reside on the domain (l, u):

$$rBeta(z \mid \alpha, \beta, l, u) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{(z-l)^{\alpha-1} \cdot (u-z)^{\beta-1}}{(u-l)^{\alpha+\beta-1}} \mathbb{1}_{\{l, u\}}(z).$$

We can use it to specify recursively a set of G conditional priors as:

$$u_1 \sim Beta(\alpha_1, \beta_1), \quad u_g|u_{g-1} \sim rBeta(\alpha_g, \beta_g; u_{g-1}, 1) \quad g = 2, \dots, G,$$

where $rBeta(\alpha_g, \beta_g; u_{g-1}, 1)$ denotes the Beta restricted to $(u_{g-1}, 1)$. The corresponding joint prior is:

$$\pi(u_1, u_2) = \frac{\Gamma(\alpha_1 + \beta_1)\Gamma(\alpha_2 + \beta_2)}{\Gamma(\alpha_1)\Gamma(\beta_1)\Gamma(\alpha_2)\Gamma(\beta_2)} \cdot u_1^{\alpha_1 - 1}(1 - u_1)^{\beta_1 - \beta_2 - \alpha_2}(u_2 - u_1)^{\alpha_2 - 1}(1 - u_2)^{\beta_2 - 1} \mathbb{1}_{\{u_1, 1\}}(u_2),$$

which does not allow for an analytical marginalization to get $\pi(u_2)$ in the general scenario. However, the *rBeta* expected value and variance are available in closed form, and hence, we can use the *law of total expectation* to derive the marginal expected value and variance of all the components. For $g = 2, \ldots, G$:

$$\mathbb{E}[u_g] = \frac{\alpha_g}{\alpha_g + \beta_g} + \mu_{g-1}\frac{\beta_g}{\alpha_g + \beta_g}$$

$$\mathbb{V}[u_g] = \sigma_{g-1}^2 \cdot \frac{\beta_g^2}{(\alpha_g + \beta_g)^2}\left(1 + \frac{\alpha_g}{\beta_g(\alpha_g + \beta_g + 1)}\right) + (1 - \mu_{g-1})^2 \cdot \frac{\alpha_g \beta_g}{(\alpha_g + \beta_g)^2(\alpha_g + \beta_g + 1)},$$

where $\mu_{g-1} = \mathbb{E}[u_{g-1}]$ and $\sigma_{g-1}^2 = \mathbb{V}[u_{g-1}]$. Therefore, one can define a system of equations to find the combination of α_g, β_g that complies with a prior knowledge on the moments of the parameters (see Figure B1b as an example).

A convenient parameter setting. A convenient parameter setting arises when $G = 2$ and if we fix $\alpha_1 = \alpha, \alpha_2 = 1, \beta_1 = \beta + 1$, and $\beta_2 = \beta$. Indeed, if $u_1 \sim Beta(\alpha, \beta + 1)$ and $u_2 \mid u_1 \sim rBeta(1, \beta, u_1, 1)$, then the joint distribution of (u_1, u_2) is:

$$\pi(u_1, u_1) = \frac{(\alpha + \beta)\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot u_1^{\alpha - 1}(1 - u_2)^{\beta}.$$

Marginalizing with respect to u_1, we get:

$$\pi(u_2) = \int_0^1 \pi(u_1, u_2) \mathbb{1}_{\{u_1, 1\}}(u_2) \, du_1 = \int_0^1 \pi(u_1, u_2) \mathbb{1}_{\{0, u_2\}}(u_1) \, du_1 = \int_0^{u_2} \pi(u_1, u_2) \, du_1 =$$

$$= \frac{(\alpha + \beta)\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot (1 - u_2)^{\beta - 1} \cdot \int_0^{u_2} u_1^{\alpha - 1} \, du_1 =$$

$$= \frac{\Gamma(\alpha + \beta + 1)}{\Gamma(\alpha)\Gamma(\beta)} \cdot (1 - u_2)^{\beta - 1} \cdot \frac{u_2^{\alpha}}{\alpha} =$$

$$= \frac{\Gamma(\alpha + \beta + 1)}{\Gamma(\alpha + 1)\Gamma(\beta)} \cdot u_2^{(\alpha + 1) - 1}(1 - u_2)^{\beta - 1},$$

which is a standard Beta density $\pi(u_2) = Beta(u_2 \mid \alpha + 1, \beta)$. Expected value and variance can then be derived from basic properties of the Beta distribution.

APPENDIX C: MODELING THE TEMPORARY EMIGRATION

Here, we show that the introduction of the *undetectability parameter* δ on the part-time individuals is equivalent to allowing for random temporary emigration. We focus on the higher hierarchy level of the part-time model specification, as all other components are not affected. For the sake of clarity, we drop the g-subscript and let the reference to the part-time group be implied.

The proposed model specification for the part-time group is as follows:

$$y_{it} \mid z_{it} \sim Bern(p_t \cdot (1 - \delta) \cdot z_{it}),$$

$$z_{it} \mid z_{i,t-1}, r_{it} \sim Bern(\phi_t \cdot z_{i,t-1} + \rho_t \cdot r_{it}), \quad r_{it} = \min\{r_{i,t-1}, 1 - z_{i,t-1}\}, \tag{C1}$$

which is different from the other groups only through the introduction of the parameter $\delta \in (0, 1)$ in the detection process.

Let us recall that z_{it} is a latent variable indicating whether individual i is *"alive"* at time t. This is generally confounded with permanent emigration, but it cannot account for temporary emigration as exited individuals cannot ever reenter the study area and return susceptible to captures. Therefore, this first latent variable is only able to model the time at which individual i starts visiting the area (is born) and the time at which it stops visiting it for good (dies). The explicit modeling of temporary emigration within this time window requires the introduction of an additional latent variable v_{it} denoting whether individual i is present given that $z_{it} = 1$. The specification of Equation (C1) arises if we assume that temporary emigration occurs randomly and with equal probability δ while individual i is alive. This corresponds to the following hierarchical specification:

$$y_{it} \mid v_{it} \ \sim \ Bern(p_t \cdot v_{it}),$$
$$v_{it} \mid z_{i,t} \ \sim \ Bern((1 - \delta) \cdot z_{i,t}), \qquad (C2)$$
$$z_{it} \mid z_{i,t-1}, r_{it} \ \sim \ Bern(\phi_t \cdot z_{i,t-1} + \rho_t \cdot r_{it}), \quad r_{it} = \min\{r_{i,t-1}, 1 - z_{i,t-1}\},$$

where $y_{it} \perp z_{it}$ if v_{it} is known. We can easily marginalize v_{it} out of Equation (C2) by noting that each $y_{it}|z_{it}$ is a Bernoulli random variable with probability of success:

$$\tilde{p}_t = \mathbb{E}[y_{it} \mid z_{it}] = \mathbb{E}[\mathbb{E}[y_{it} \mid v_{it}] \mid z_{it}] =$$
$$= \mathbb{E}[p_t \cdot v_{it} \mid z_{it}] = p_t \cdot \mathbb{E}[v_{it} \mid z_{it}] = p_t \cdot (1 - \delta) \cdot z_{it},$$

from which:

$$y_{it} \mid z_{it} \ \sim \ Bern(p_t \cdot (1 - \delta) \cdot z_{it}).$$

This equivalence and the corresponding interpretation are what motivates the use of a multiplicative parametrization on the capture probability at the invlogit scale in place of a more straightforward group-specific intercept within the logit specification.

APPENDIX D: DETAILS ABOUT TIME LAGS USED IN THE SIMULATION STUDY

The number of days between two consecutive capture occasions (*daily time lags*, henceforth) within a single year has been simulated from a *shifted* geometric distribution with probability 0.05, which has an expected value equal to 20 and a standard deviation equal to 19.5. The resulting random sequence of time lags is

$$(20, 1, 12, 15, 56, 9, 9, 12, 10),$$

and, for scenarios that contemplate more than 1 year of study, the same sequence is repeated during each new year. The shift of year occurring each 10 occasions is achieved by using a higher constant time lag (i.e., 240 days) between the $(10k)$th occasion and the $(10k + 1)$th occasion, with $k = 1, 2, 3$. This results in a scenario k composed by k years of study, for $k = 1, 2, 3, 4$. For example, scenario 2 ($T = 20$) is composed of the following sequence of time lags, resulting in 2 years of capture occasions:

$$(20, 1, 12, 15, 56, 9, 9, 12, 10, 240, 20, 1, 12, 15, 56, 9, 9, 12, 10).$$

APPENDIX E: CONVERGENCE OF RELEVANT PARAMETERS ESTIMATED ON REAL DATA

We checked the convergence of the parameter chains in the real data application through the general-purpose Gelman diagnostic. All potential-scale reduction factors $\hat{R}$ are below 1.01, which suggests a good mixing of all parameter chains. In Figure E1, we show the behavior of the traceplots of the most relevant parameters. We can observe how the two chains explore the same parameters space in all cases and produce a well-shaped posterior distribution, with no bad behavior. The corresponding density estimates are shown in Figure E2.
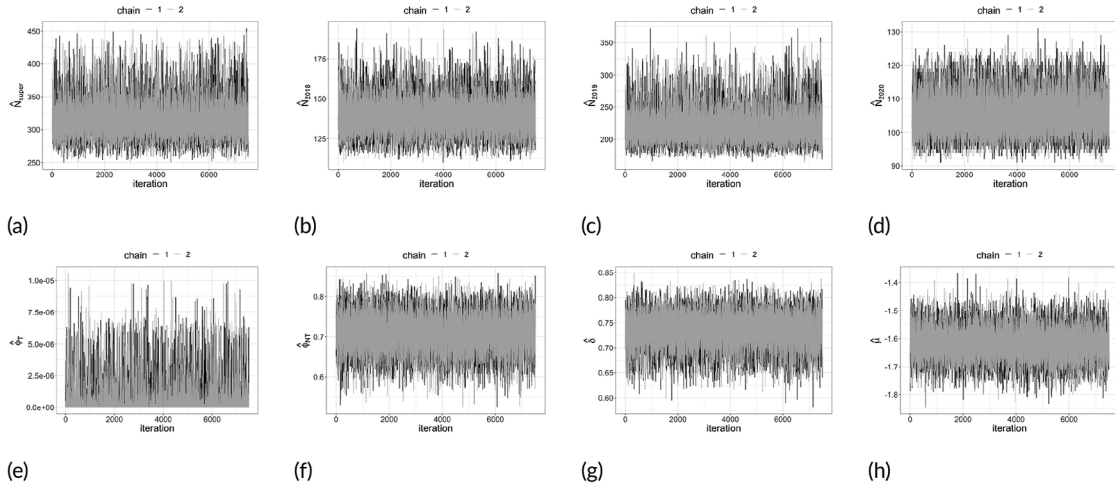
FIGURE E1 Traceplots of posterior samples for the main parameters of interest of model RPT.
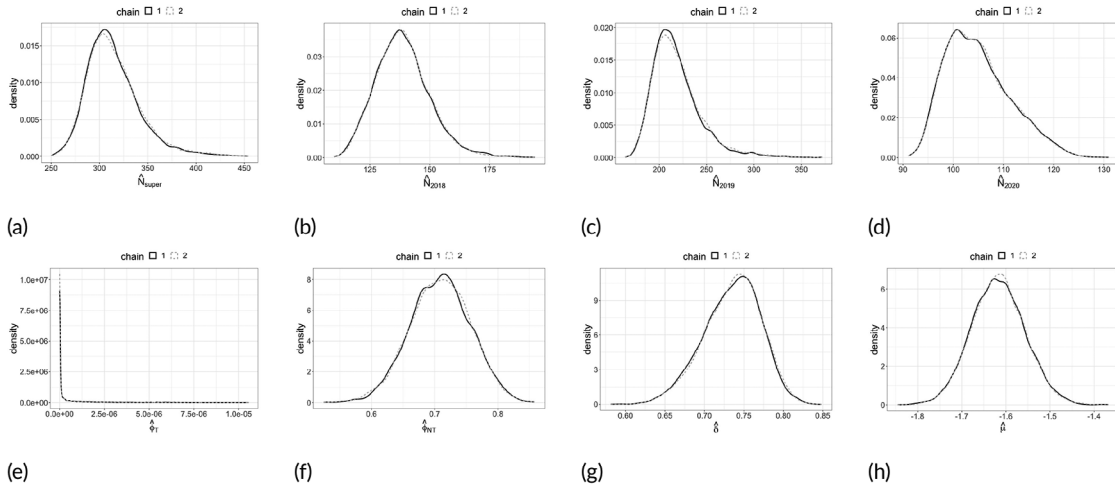

FIGURE E2 Densities of posterior samples for the main parameters of interest of model RPT.

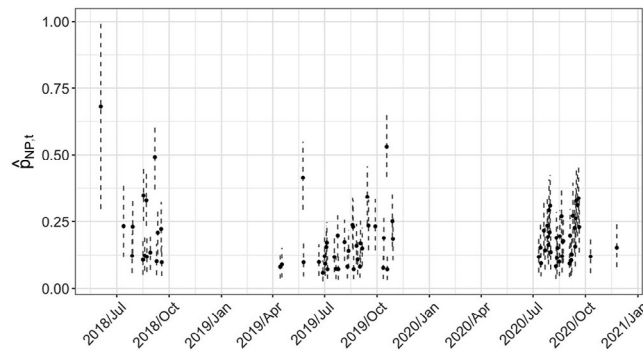## APPENDIX F: ESTIMATED TIME-VARYING PARAMETERS OF THE RPT MODEL ON THE REAL DATA APPLICATION

FIGURE F1 Posterior estimates and 95% credible intervals for capture probabilities of resident and transient individuals at each sampling occasion.
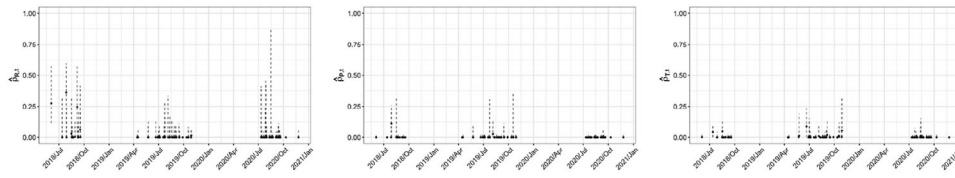
FIGURE F2 Estimated recruitment probabilities (a) $\rho_{R,t}$, (b) $\rho_{P,t}$, and (c) $\rho_{T,t}$.