Submitted to the Annals of Applied Statistics
2005 Vol. 0, No. 0, 1–8
arXiv: arXiv:0000.0000

MODELLING INDIVIDUAL MIGRATION PATTERNS USING A BAYESIAN NONPARAMETRIC APPROACH FOR CAPTURE-RECAPTURE DATA

By Eleni Matechou and François Caron

University of Kent and University of Oxford

We present a Bayesian nonparametric approach for modelling wildlife migration patterns using capture-recapture (CR) data. Arrival times of individuals are modelled in continuous time and assumed to be drawn from a Poisson process with unknown intensity function, which is modelled via a flexible nonparametric mixture model. The proposed CR framework allows us to estimate: i) the total number of individuals that arrived at the site, ii) their times of arrival and departure and hence their stopover duration, and, iii) the density of arrival times, providing a smooth representation of the arrival pattern of the individuals at the site. We apply the model to data on breeding great crested newts (*Triturus cristatus*) and on migrating reed warblers (*Acrocephalus scirpaceus*). For the former, the results demonstrate the staggered arrival of individuals at the breeding ponds and suggest that males tend to arrive earlier than females. For the latter, they demonstrate the arrival of migrating flocks at the stopover site and highlight the considerable difference in stopover duration between caught and not-caught individuals.

1. Introduction. Many wildlife populations migrate between their overwintering sites and breeding sites twice a year. This is especially true for populations of birds but also mammals and amphibians. In recent years, several species have been observed to change their phenology with populations spending less time at their overwintering sites and moving earlier to their breeding sites than in the past. These changes are mostly attributed to the warming climate (see for example Bauer et al., 2008; Van Buskirk et al., 2009; Sullivan et al., 2015). We note here that phenology is defined by the Oxford English dictionary as "The study of cyclic and seasonal natural phenomena, especially in relation to climate and plant and animal life" and hence we use the term to refer to migration and breeding patterns, which are of course interlinked.

As Seebacher and Post (2015) state "...(the) global geographical scale (of migration) makes migrating individuals particularly vulnerable to climate change, and at the same time, the process of migration has fundamental impacts on ecological processes and biodiversity". According to Both et al. (2009), changes in climate have in some cases led to a

MSC 2010 subject classifications: Primary 60K35, 60K35; secondary 60K35
Keywords and phrases: Chinese restaurant process, great crested newts, Poisson-Gamma process, reed warblers, shot-noise Cox process, stopover data

mismatch between the peak food availability and phenology which has resulted in declines of numbers in some species.

Hence, it is crucial to monitor phenology of populations, as well as the duration of time that individuals spend at the site(s), termed stopover duration, and population sizes. This information can be useful in assessing for example the importance of a particular site or in informing about the effect of, or need for, conservation strategies.

The work in this paper is motivated by capture-recapture (CR) data, such as the data represented in Fig. 2 (a), that are often collected at sites of interest. CR data result from repeatedly sampling a population and uniquely marking newly caught individuals before releasing them back into the population. We consider two case studies:

i) CR data on great crested newts (GCN) (*Triturus cristatus*) collected in the UK (Section 3.1). GCN are a European protected species. They overwinter away from water and in late winter they migrate to ponds in order to breed, their phenology influenced by weather conditions (Lewis, 2012);
ii) CR data on reed warblers (RW) (*Acrocephalus scirpaceus*) collected in Switzerland (Section 3.2). RW overwinter in Africa and migrate to Europe by travelling short distances at a time and utilising stopover sites along the way. Kovács et al. (2012) reported that in Hungary spring migration of RW has in recent years shifted a week earlier while autumn migration a week later, agreeing with patterns reported for migrating species in general.

CR data can be analysed using Jolly-Seber (JS) type models (Jolly, 1965; Seber, 1965; Schwarz and Arnason, 1996; Pledger et al., 2009; Matechou et al., 2013b) which account for the sampling scheme and for new individuals arriving into the population, as well as for individuals leaving the population (Cormack, 1964; Lebreton et al., 1992). Typically fitted using a frequentist approach, JS models are not built at the individual level so as to avoid dealing with a large number of latent variables, which is challenging. Instead, they are built at the population level and hence estimate the proportion of individuals that were new arrivals at each sampling occasion, instead of individual arrival times. Any inference drawn is restricted to the population as a whole. But as Charmantier and Gienapp (2014) note, it is the information at the individual level that will allow us to study and understand any changes in phenology. Additionally, arrival is modelled in discrete time and the total number of individuals that became available for detection at least once, termed the "super-population", is estimated instead of the total number of individuals that visited the site. Hence, inference does not account for individuals that had shorter stopover durations and departed before ever becoming available for detection. Therefore, the population size estimates obtained by these models can be different depending on the length of the intervals between sampling occasions, an undesirable feature similar to the issue of length-biased sampling in survival analysis.

More recently, Bayesian formulations of the JS model have also been considered, as in Royle et al. (2007) and Lyons et al. (2015). These can be used to estimate individual arrival

times but they still model arrival in discrete time and hence share some of the limitations of their frequentist predecessors. In addition, since the population size is unknown and possibly updated at each iteration of the algorithm used to fit the model, trans-dimensional algorithms, such as reversible jump Markov chain Monte Carlo (MCMC) (Green, 1995) or data-augmentation techniques (Royle and Young, 2008) are employed to deal with the changing dimensions of the model. However, the former can be difficult to set up and tune and the latter requires the specification of an upper bound for the population size, which is not typically known.

In this paper we adopt a Bayesian nonparametric approach for modelling the arrival of individuals into the population in continuous time using a flexible mixture model. We propose a CR model which allows us to estimate the total number of individuals that visited the site and to reconstruct the unknown presence histories of individuals, i.e. to estimate their times of arrival and departure, and hence the total amount of time they spent at the site. This allows us to compare estimated arrival times and stopover duration between individuals that were eventually caught and those that were never caught, as well as between individuals with different characteristics, such as sex. Additionally, the use of our proposed mixture model to represent the arrival pattern enables us to overcome the issue of length-biased sampling mentioned above since individuals that arrived at the site but never became available for capture are also accounted for in the population. Finally, we propose an elegant MCMC update for the population size using forward simulation from the model which is an alternative to data-augmentation techniques commonly employed in similar models.

We treat the data as generated by a marked Poisson process which consists of three parts: the arrival, departure and capture processes. For the arrival process, we model the unknown arrival times of individuals as a shot-noise Cox process (Wolpert and Ickstadt, 1998; Brix, 1999; Møller, 2003). More precisely, arrival times are assumed to be drawn from a Poisson process whose intensity is itself random, and modelled by a mixture (Lo and Weng, 1989; Kuo and Ghosh, 1997; Nieto-Barajas and Walker, 2004; Ishwaran and James, 2004; Kottas and Sansó, 2007; Taddy and Kottas, 2012). We allow for an unknown number of mixing components and to accommodate them, we assume that the intensity takes the form of an infinite mixture of normal distributions, whose mixing distribution is a gamma process (Wolpert and Ickstadt, 1998; Brix, 1999). We derive an MCMC sampler for posterior inference on the size of the population as well as on the arrival and departure times of individuals; importantly, due to the analytic properties of our Bayesian nonparametric model, the sampler does not require designing explicit trans-dimensional moves (Green, 1995), which, as mentioned above, may be difficult to tune.

The marked Poisson process model for CR data is described in Section 2 with more details about the gamma process given in Appendix A. The hierarchical representation of the model is given in Appendix B and an MCMC algorithm for posterior inference in Appendix C. The two case studies are presented in Section 3 and a comparison of the results to those obtained by an existing JS type model is presented in the supplementary

material.

2. Model.

2.1. *Data.* Data are collected at the defined study site on K sampling occasions, which are assumed to be instantaneous, taking place at times $t_1 < t_2 \ldots < t_K$ and indexed by $k = 1, \ldots, K$. On each of these sampling occasions, the population is sampled appropriately, for example using nets or traps, and all caught individuals are uniquely marked, unless they were already marked, and then released back into the population.

Let N be the unknown population size and $D \leq N$ the observed number of individuals caught at least once. We use i to index individuals with $i = 1, \ldots, N$. We denote by $\mathrm{H}_i \in \{0,1\}^K$ the capture history of individual i, with an entry of 1 denoting that individual i was caught on that particular sampling occasion and 0 otherwise. The $N - D$ individuals that were never caught share the capture history with all entries equal to 0. The data set $\mathcal{D}$ consists of all the D capture histories with at least one non-0 entry. We note here that N does not correspond to the total number of individuals that became available for capture at least once during the study, which is the definition of the "super-population" size in for example Schwarz and Arnason (1996) and other JS-type models. Instead, in our case individuals that arrived and departed without ever becoming available for capture are also accounted for.

For an example of a CR data set the reader is directed to Figure 2 (a).

2.2. *Marked Poisson process.* Each of the N individuals entered and exited the study site during one of the

$$T_0 = (-\infty, t_1), T_1 = [t_1, t_2), \ldots, T_K = [t_K, +\infty)$$

intervals. Note that if an individual exited in T_0 or entered in T_K, or entered and exited in the same interval then it never became available for capture. Individuals that were already present at the start of the study entered in T_0 while individuals that were still present after the end of the study exited in T_K.

The arrival time of individual i is denoted by ζ_i, with $\zeta_i \in \mathbb{R}$. We denote by $b_i \in \{0, 1, \ldots, K\}$ the index of the interval in which individual i entered the population and by d_i the index of the interval in which it departed, with $b_i \leq d_i \in \{0, \ldots, K\}$.

We consider that the points $\{(\zeta_i, d_i, \mathrm{H}_i)\}_{i=1,\ldots,N}$ are the points of a marked Poisson process (Kingman, 1993; Daley and Vere-Jones, 2008). Specifically, the arrival times $(\zeta_i)_{i=1,\ldots,N}$ are drawn from a non-homogeneous Poisson process of intensity $\nu(\zeta|G)$ and for $i = 1, \ldots, N$ the marks (departure d_i and capture H_i) are generated from

$$d_i|\zeta_i \sim \Pr(d_i|\zeta_i, \gamma)$$
$$\mathrm{H}_i\,|\zeta_i, d_i, \beta \sim \Pr(\mathrm{H}_i\,|\zeta_i, d_i, \beta),$$

where (G, β, γ) is a set of hyperparameters. We present the details on the arrival, departure and capture processes in the following sections.

2.3. *Arrival process.* The unknown intensity function ν tunes the arrival pattern of the individuals at the study site. Note that the Poisson process construction implies that the population size N is drawn from a Poisson distribution with rate $\omega = \int_{-\infty}^{\infty} \nu(\zeta|G)d\zeta$, the overall intensity level.

Arrivals of migrating individuals tend to be synchronised, with individuals either travelling together towards specific sites or anyway arriving in a synchronised manner because their migration is triggered by common environmental or individual factors. Hence, we assume that individuals become part of the population by entering the study site in clusters, which can potentially overlap in their arrival times. Specifically, we consider that the positive intensity function ν takes the form of a mixture of normal distributions, parametrized by an unknown mixing distribution G, which is an (unnormalized) random measure,

$$\nu(\zeta|G) = \int_{-\infty}^{\infty} \int_{0}^{\infty} \mathcal{N}(\zeta; \mu, \sigma^2) G(d\mu, d\sigma^2) \tag{2.1}$$

where $\mathcal{N}(\zeta; \mu, \sigma^2)$ denotes the probability density function (pdf) of a normal random variable with mean μ and variance σ^2 evaluated at ζ. The choice of a normal pdf for representing the arrival pattern leads to an efficient MCMC algorithm and allows us to unearth the major patterns in the arrival process.

We adopt a Bayesian nonparametric approach and assume that G is infinite-dimensional, drawn from a gamma process (Kingman, 1993). The gamma process is parametrized by two parameters $\alpha > 0$, $\tau > 0$ and a probability measure G_0. Parameters α and τ both tune the overall intensity level, ω, with $\omega \sim \text{Gamma}(\alpha, \tau)$, where $\text{Gamma}(a, b)$ denotes the standard gamma distribution of shape $a > 0$ and inverse scale $b > 0$. α also tunes the variability of the relative sizes of the different clusters, with lower values corresponding to higher variability. Note that the overall intensity ω, and thus the population size N are both almost surely finite.

Parameter G_0 is a prior distribution on the means, μ, and variances, σ^2, of the arrival times of each cluster. For computational convenience, we set G_0 to be a normal inverse gamma distribution, which is a conjugate prior for the normal distribution: $(\mu, \sigma^2) \sim G_0$ stands for

$$\mu|\sigma^2 \sim \mathcal{N}(m_0, \sigma^2/\kappa_0) \tag{2.2}$$

$$1/\sigma^2 \sim \text{Gamma}(\nu_0, \lambda_0). \tag{2.3}$$

where $m_0 \in \mathbb{R}$, $\kappa_0 > 0$, $\lambda_0 > 0$ and $\nu_0 > 0$ are tuning parameters.

The parameters α and τ are themselves considered to be unknown, with

$$\alpha \sim \text{Gamma}(a_\alpha, b_\alpha), \qquad \tau \sim \text{Gamma}(a_\tau, b_\tau).$$

Details on the setting of the hyperparameters for the applications considered in this paper are given in Section 2.7 while more details on the gamma process are given in Appendix A.

2.4. *Departure process.* We assume that each individual i departs from the study site with a piecewise constant hazard rate $\Lambda(t)$

$$\Lambda(t) = \sum_{k=0}^{K} \lambda_k \mathbb{1}_{T_k}(t), \tag{2.4}$$

where $\mathbb{1}_A(t) = 1$ if $t \in A$ and 0 otherwise and, for $k = 0, \ldots, K$,

$$\lambda_k = \log\left(1 + \exp(-x_k^\intercal \gamma)\right) \tag{2.5}$$

where $x_k \in \mathbb{R}^q$ is a vector of covariate values associated to interval $k = 0, \ldots, K$ and $\gamma \in \mathbb{R}^q$ is a vector of coefficients with $\gamma \sim \mathcal{N}(0_{\widetilde{q}}, I_{\widetilde{q}})$. Hence, given arrival time $\zeta_i \in T_{b_i}$, the probability that individual i departs in interval d_i is

$$\begin{aligned}
\Pr(d_i|\zeta_i, \gamma) &= e^{-\int_{\zeta_i}^{t_{d_i}} \Lambda(t)dt} \left(1 - e^{-\int_{t_{d_i}}^{t_{d_i+1}} \Lambda(t)dt}\right) \\
&= \begin{cases} e^{-(t_{b_i+1}-\zeta_i)\lambda_{b_i}} \left[\prod_{k=b_i+1}^{d_i-1} e^{-(t_{k+1}-t_k)\lambda_k}\right] \left[1 - e^{-(t_{d_i+1}-t_{d_i})\lambda_{d_i}}\right] & \text{if } d_i > b_i \\ 1 - e^{-(t_{d_i+1}-\zeta_i)\lambda_{d_i}} & \text{if } d_i = b_i \end{cases}
\end{aligned}$$

Defining $\phi_k^{(t_{k+1}-t_k)} = e^{-(t_{k+1}-t_k)\lambda_k}$ as the probability of surviving from time t_k to time t_{k+1} we obtain:

$$\Pr(d_i|\zeta_i, \gamma) = \begin{cases} \phi_{b_i}^{(t_{b_i+1}-\zeta_i)} \left\{\prod_{k=b_i+1}^{d_i-1} \phi_k^{(t_{k+1}-t_k)}\right\} \left\{1 - \phi_{d_i}^{(t_{d_i+1}-t_{d_i})}\right\} & \text{if } d_i > b_i \\ 1 - \phi_{d_i}^{(t_{d_i+1}-\zeta_i)} & \text{if } d_i = b_i \end{cases} \tag{2.6}$$

We note that this expression is similar to those used in JS-type models, such as the Pledger et al. (2009) model, and it allows us to consider a range of parameterisations for ϕ, which can be for example considered to be constant for the duration of the study or dependent on time-varying covariates.

2.5. *Capture process.* If $\widetilde{x}_k \in \mathbb{R}^{\widetilde{q}}$ is a vector of covariate values on sampling occasion $k = 1, \ldots, K$ and $\beta \in \mathbb{R}^{\widetilde{q}}$ is a vector of coefficients with $\beta \sim \mathcal{N}(0_q, I_q)$ then the probability that the k^{th} entry H_{ik} of the observed capture history H_i is equal to 1, i.e. the probability that individual i was caught on sampling occasion k, is

$$\Pr(\mathrm{H}_{ik} = 1|\zeta_i, d_i, \beta) = \begin{cases} \frac{1}{1+\exp(-\widetilde{x}_k^\intercal \beta)} & \text{if } b_i < k \leq d_i, \\ 0 & \text{otherwise.} \end{cases}$$

We note here that we have chosen the prior variance-covariance matrices for both γ and β to be the identity matrices but we show in our supplementary material that our inference is not affected by the choice of prior in this case since we obtain the same posterior distributions for these parameters when we specify the diagonal of these matrices to be 10^2 or even 100^2.

2.6. *Model fitting.* The hierarchical representation of the whole model is given in Appendix B and details on the MCMC algorithm for posterior inference on the model parameters are given in Appendix C. The accompanying R code (R Core Team, 2014) is available at *link to online supplementary material.*

For both applications considered in this paper, we run three chains of the algorithm, using starting values for the parameters randomly generated from the parameter space. We discarded 50000 iterations and thinned the chains by keeping one every 300 samples. We concluded convergence by visual inspection of trace plots and by the Gelman-Rubin diagnostic plot produced using the R-package coda (Plummer et al., 2006). These diagnostics are presented in the supplementary material.

2.7. *Hyperparameter settings.* The parameters of G_0 have to reflect our prior beliefs and understanding about the arrival process of the population. We expect the arrival times of clusters to be mostly within the study limits, by study design, as the populations are non-resident and the sampling period is expected to encompass the residency period. Hence we have chosen the parameters of G_0 to reflect that, while also allowing for values outside that range to be proposed with a lower frequency. The arrival times of each cluster are not expected to span more than a few sampling occasions, with clusters potentially arriving in short, abrupt bursts. We chose to set $\nu_0 = 4$ and $\lambda_0 = 1$ so that 95% of the distribution mass for the standard deviation of arrival times is between 0.3 and 0.96. This prior is flexible enough to allow for the creation of clusters with arrival times which span anything between one and a few (eg. four) sampling occasions. We set $\mu_0 = \tau_K/2$ and chose the value for κ_0 so that, a priory, roughly 95% of the arrival times simulated from G_0 fall within the study limits, which is expected when studying migrating populations. In particular, for the example shown in Section 3.1 we set $\kappa_0 = 0.01$ while for the example in Section 3.2 we set $\kappa_0 = 0.03$. Finally, we choose improper priors for parameters α and τ and set $a_\alpha = b_\alpha = a_\tau = b_\tau = 0$.

3. Applications.

3.1. *Great crested newts.* The data set, collected by the Durrell Institute of Conservation and Ecology, University of Kent, concerns a small population of GCN that breeds in eight artificial ponds that are located on the university campus (Lewis, 2012). GCN hibernate on land and migrate to ponds in spring in order to breed. Once their breeding is complete, they return to land to overwinter. Individual GCN are uniquely identifiable

by their belly patterns and male GCN are distinguished from females by the crest on their backs. During the breeding season of 2012, $D = 30$ adult GCN were caught at least once in $K = 22$ weekly sampling occasions. Here, $t_1 = 1$ and $t_k - t_{k-1} = 1$, $k = 2, \ldots, K$.

We assume that capture probability is a function of the number of traps placed in the ponds, which is either 6 or 8, and that survival probability varies by calendar time, as all of the GCN will leave the ponds by the end of the breeding season, and we use a logistic regression model with standardised week number, 1-22, as the covariate to represent that dependence.

We estimate that the probability that all of the GCN present that season were caught is less than 20%, while the probability that more than 5 GCN were missed is $\approx 5\%$ (Figure 1 (a)). The posterior mean for capture probability is equal to 0.40 with (0.30, 0.49) 95% posterior credible interval (PCI) when the number of traps is 8. This is similar to summaries obtained when the number of traps is 6 (mean = 0.39, 95% PCI = (0.33, 0.46)), which is due to the fact that the number of individuals that can be caught each week is not limited by the number of traps. Note that the PCI around capture probability in the second case is marginally narrower since more samples where collected using 6 rather than 8 traps.

Figure 1 (b) plots posterior draws of the normalized intensity, or density of the arrival times ζ at 500 randomly chosen iterations of the algorithm, shown by the gray lines, as well as the posterior mean normalized intensity, shown by the black line. The mean normalized intensity for ζ provides a smooth representation of the arrival pattern of the GCN at the breeding site and suggests an almost continuous flow of arriving individuals, at least for the first half of the season. The boxes at the bottom of the plot represent the values of ζ that fall in the 95% highest posterior density (HPD) interval, constructed using R package "hdrcde". The figure suggests that a high proportion of GCN were already present at the start of the study (roughly 47%). Almost 95% of GCN are estimated to have arrived by week 12. Weeks 2 and 15-22 are outside the 95% HPD interval of arrival times, suggesting possibly two major arrival groups with migration to the ponds concluding by roughly the middle of the season.

The estimates of individual arrival times of GCN caught at least once suggest that male GCN arrive at the breeding ponds earlier than females, agreeing with the literature on the ecology of the species (Jehle et al., 2011). Specifically, almost 60% of caught males are estimated to have been present when the study commenced, while the corresponding proportion for females is around 10%. Additionally, males are estimated to be present at the start of the study on average while females arrive much later, on average between weeks 5 and 6.

As expected, survival probability is estimated to decrease considerably by week (Figure 1 (c)). The 95% PCI for d includes weeks 4-22 and has posterior mean equal to 15.5 with only around 1% of GCN estimated to still be present at the end of the study period, i.e. with $d \geq 22$.

Finally, to check the fit of the model, we generated CR data from the reconstructed presence histories obtained at a random sample of iterations of the algorithm, and plotted

the observed number of individuals caught each week together with the means and 95% percentile intervals of the simulated values (Figure 1 (d)). The model provides a satisfactory fit to the data as it is able to reconstruct the overall trend in the data, with numbers peaking around the middle of the study and gradually decreasing towards the end, as the GCN are leaving the ponds.

3.2. *Reed Warblers.* We consider the data set on migrating RW collected in a river delta in southern Switzerland and analysed by Schaub et al. (2001). Captures took place over 70 days but the data were pooled over 5-day periods, resulting in 14 capture occasions and 567 birds caught at least once. Hence, $t_1 = 1$ and $t_{k+1} - t_k = 1$, $k = 2, \ldots, K$. A representation of the data set is given in Figure 2 (a).

Schaub et al. (2001) used the recruitment approach of Pradel (1996) to estimate stopover duration before the time of first capture for each individual and standard survival analysis (Lebreton et al., 1992) to estimate stopover duration after the time of first capture. They found that recruitment was time-dependent, while survival and capture probability were constant.

Following Schaub et al. (2001), we assume that both ϕ and p are constant. The posterior means and 95% PCI for ϕ and p are found to be 0.39 (0.32, 0.45) and 0.21 (0.15, 0.29), respectively. We estimate that the population size was substantially greater than the sample size (Figure 2 (b)) with posterior mean equal to 2957 (95% PCI = (2345, 3719)). The density plot of ζ, presented in Figure 2 (c) shows that arrival times span the whole study duration with sampling occasions 3, 6, 11 and 14 outside the 95% HPD interval for ζ. The estimated arrival pattern clearly demonstrates the arrival of around four or five waves or flocks of birds at the breeding site. The synchronous arrival of migrating birds at stopover sites is the result of favourable weather, eg. wind and rain (Erni et al., 2002; Schaub et al., 2004) which is typically synchronous over large spatial scales. This results in migration waves, such as the ones shown in Figure 2 (c). Finally, the fit of the model is assessed in Figure 2 (d) using the posterior predictive distribution.

To estimate the average stopover duration, we can use the reconstructed presence histories, as obtained at each iteration of the algorithm. The proportions of the estimated difference between d and b for marked and unmarked individuals are given in Table 1. It can be seen that over 50% of marked birds are estimated to have spent at least 10 days at the site while 44% of unmarked birds have $d - b = 0$. Since the interval between sampling occasions is equal to 5 days, we use the midpoint of each interval as an approximation to the number of days birds that departed in that interval spent at the site. We note here that since we model arrival in continuous time, we could instead use the average individual estimated ζ, but this way our results for marked birds are directly comparable to those obtained by Schaub et al. (2001). For example, birds that have $d = b$ spent on average 2.5 days at the site, birds with $d - b = 1$ spent on average 7.5 days etc. The average stopover duration of caught birds is equal to 12.5 days, which is similar to the value obtained by Schaub et al. (2001) (12.3). However, the average stopover duration of unmarked birds is

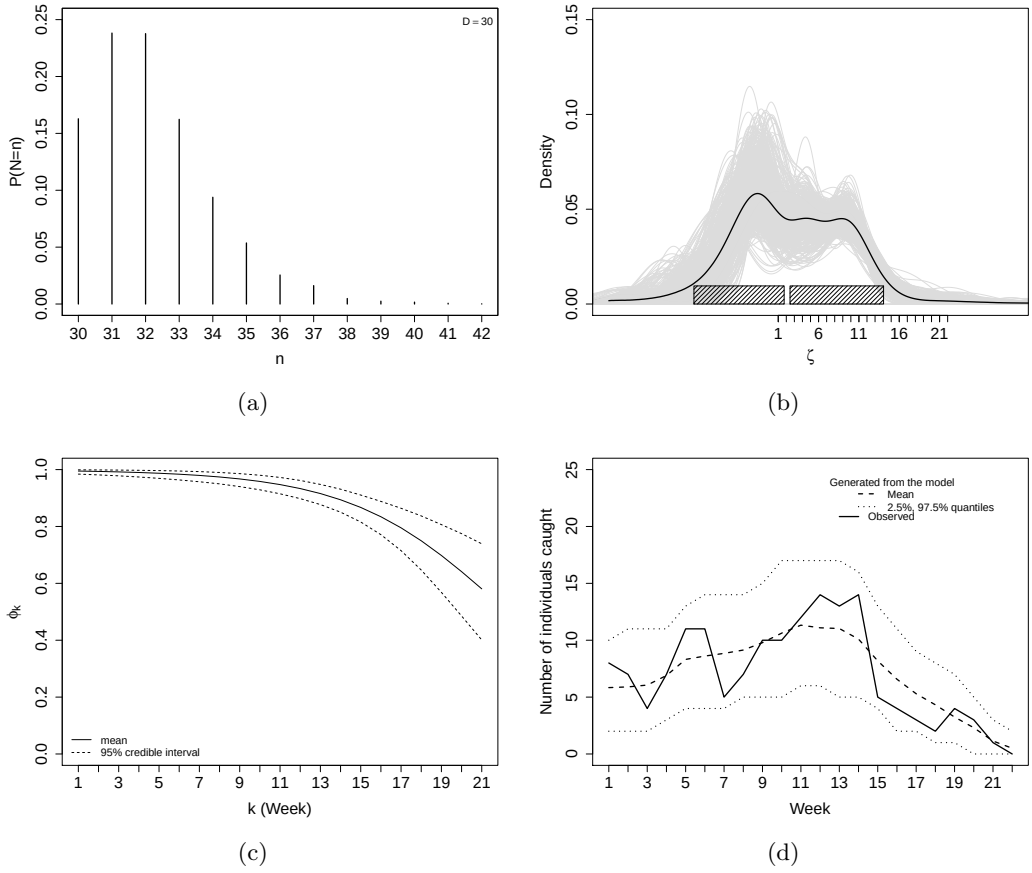

FIG. 1. *Great crested newt data. (a): Posterior distribution of N. (b): Draws from the normalized intensity of the arrival times ζ obtained at 500 randomly selected iterations of the algorithm (gray lines), with the black line showing the mean normalized intensity and the tick marks on the x-axis indicating sampling occasions. The position of the boxes on the x-axis indicates the values of ζ that fall in the 95% HPD interval while their height is equal to the lowest density value in the interval. (c): Posterior mean and 95% PCI of $\phi_k = e^{-\lambda_k}$ as a function of week number, $k = 1, \ldots, 22$. (d): Number of individuals caught each week, together with summaries of values simulated from the model.*

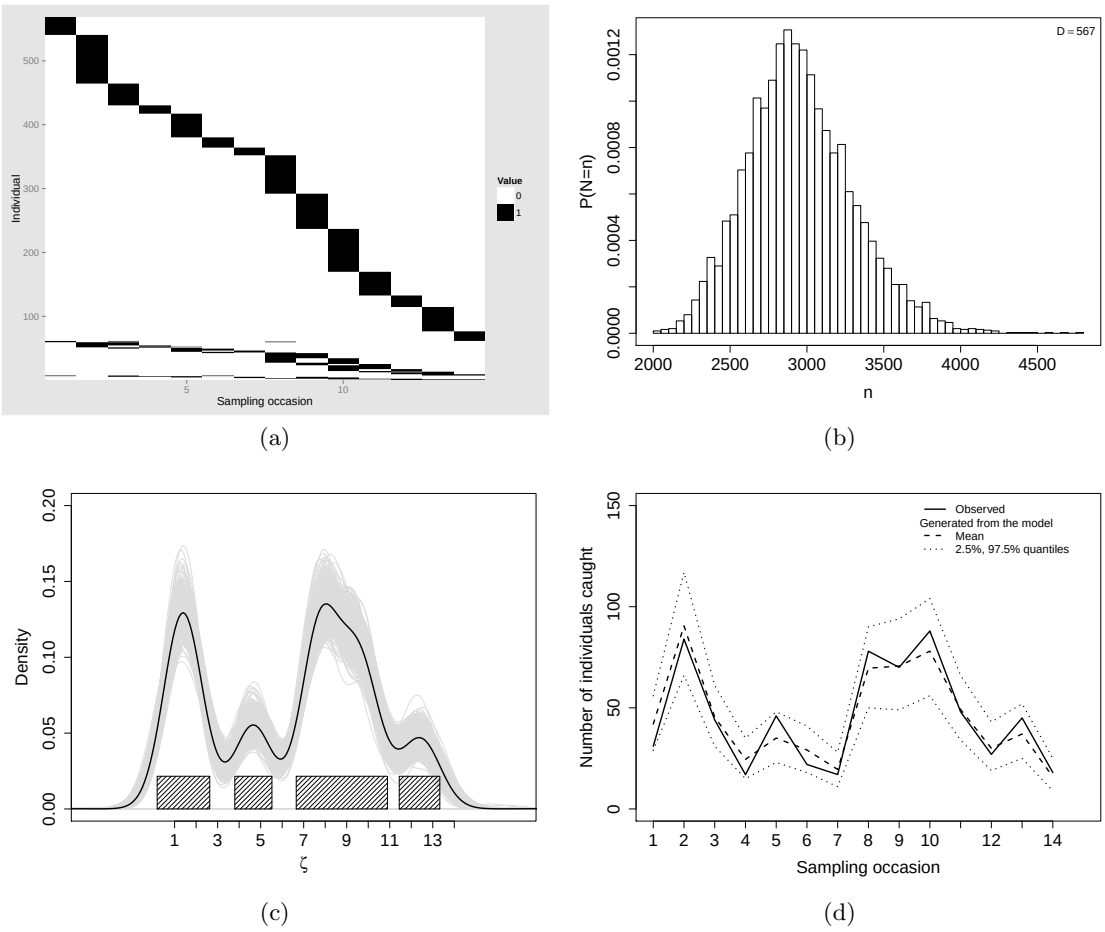

FIG. 2. *Reed warbler data. (a): Representation of the data with black blocks indicating captures and white non-captures. Individuals are ordered first by the number of times they were caught, in decreasing order, and subsequently by the time of their first capture. (b): Posterior distribution of N. (c): Density estimates for ζ obtained at 500 randomly chosen iterations of the algorithm (gray lines), with the black line showing the mean density and the tick marks on the x-axis indicating sampling occasions. The position of the boxes on the x-axis indicates the values of ζ that fall in the 95% HPD interval while their height is equal to the lowest density value in the interval. (d): Number of individuals caught each week, together with summaries of values simulated from the model.*

considerably lower (6.5 days), resulting in an overall average stopover duration of around 8 days. This difference in the stopover duration between marked and unmarked birds highlights the importance of using models, such as the one presented in this paper, that take into account the individuals that were never caught, which are likely to be the ones with the shorter stopover durations and thus overcome length-biased sampling issues.

TABLE 1
Reed warbler data. Proportion table of estimated values of $d-b$ obtained for marked and unmarked birds.

$d-b$	0	1	2	3	4	5	6
marked	0	44	30	15	7	3	1
unmarked	44	40	12	3	1	0	0

4. Discussion. In recent years, birds and other animals have been observed to change their phenology as they adapt to a changing climate. At the same time, site suitability is also changing due to increasing temperatures and other environmental changes. As a result, the distribution of wildlife populations is changing over time and space. It is crucial to monitor these adaptations and record changes in numbers or behaviours of individuals. We have presented a flexible model which provides estimates of ecologically important parameters such as population size, time spent at the site and density of arrival times, for open non-resident populations using CR data. The model can be fitted to data sets collected in different years and/or at different sites to detect any potential patterns or changes and inform about the need of policy implementation.

Our approach is an alternative to JS-type models and we present a comparison of our results to those obtained by the Pledger et al. (2009) parameterisation of the JS model in Section 3 of our supplementary material. The results between the two approaches are generally in agreement but our approach has four main advantages over the existing, frequentist and Bayesian, JS-type methods:

1. Smooth representation of the arrival pattern. By modelling arrival of individuals in continuous time, we obtain a smooth representation of the arrival pattern at the site. This is not only ecologically interesting but it can be especially useful when comparing analyses of data sets collected in different years, as potential patterns or trends over time can be detected more easily by simply comparing the posterior mean intensity function. We note here that we have treated the problem of estimating the arrival pattern as a density estimation problem and clustering of individuals arose in the process. However, these clusters can overlap, making interpretation of the number, size and other cluster characteristics challenging, hence we have not tried to interpret them from an ecological perspective.

2. Overcoming the issue of length-biased sampling. Our model allows us to estimate the total number of individuals that arrived at the site as opposed to the number that became available for detection at least once. As our results in Section 3 of our supplementary material demonstrate, these two values can be considerably different if the intervals between

sampling occasions are long compared to the average stopover duration of individuals in the population. In addition, individuals that arrived but departed before the start of the study are also accounted for, which is not the case in for example Lyons et al. (2015) who mention that, typically, studies at stopover sites are planned so that they start before most individuals have arrived. However, since phenology is changing in recent years, satisfying this criterion can become increasingly more difficult. This kind of bias is often encountered in ecological applications where detection is imperfect. For example, Gilbert et al. (2014) state that "Estimates of survival from neonates that are opportunistically captured might be inaccurate because some individuals die before sampling, resulting in data that are left truncated.". Hence, our approach could be modified for modelling time of birth instead of time of arrival to account for individuals that never became available for detection and correct such bias.

3. Estimation of individual arrival/departure times. Since we are estimating individual arrival and departure times, similarly to Lyons et al. (2015), we are able to estimate individual stopover durations as well as other statistics that are potentially of interest, such as number of individuals present at any time point. However, in contrast to Lyons et al. (2015), we do not assign an arrival time of one to all individuals that were already present at the start of the study as our mixture model allows us to extend arrival to times prior to the start of the study, while accounting for the probability of remaining at the site until the start of the study. For the applications considered in this paper, estimated individual arrival and departure times allowed us to compare the arrival pattern of individuals of different sex as well as the estimated stopover duration of individuals that were caught at least once with that of individuals that were never caught. Additionally, if data for multiple years are available, our model enables monitoring the arrival times of specific individuals over different years, and potentially linking them to other ecological processes of interest.

4. Estimation of N. When updating the population size, N, the parameter vector dimension also changes. However, our approach for estimating the size of the population does not require the use of reversible jump MCMC algorithms, or the specification of an upper bound, as in data-augmentation techniques, to perform this update. Our proposed framework is general and it can be applied to other similar models when only a subset of the population is observed and updates of N are performed as part of the estimation process. A similar, and topical, application where data-augmentation has been considered is in the area of spatially-explicit CR models (see Royle et al., 2009, for example) where the probability of detecting an individual is a function of the (unknown) distance of the trap from the centre of its home-range.

We have chosen to model the unknown intensity of arrivals as a shot-noise gamma process. There is a rather large literature on Cox processes, see e.g. (Møller and Waagepetersen, 2004, Chapter 5). A standard alternative is the log-Gaussian Cox process, where the log-intensity is drawn from a Gaussian process (Møller et al., 1998; Brix and Diggle, 2001). The approach we have chosen has however a number of advantages over the log-Gaussian Cox process: i) it directly provides a prior over continuous intensity functions, without the

need for a transformation, ii) it can naturally capture multiple modes, corresponding to the arrival pattern of different arrival groups and iii) it ensures that the overall intensity ω is finite almost surely, and its (gamma) distribution is explicitly known; in the log-Gaussian process case, this intensity may be infinite; even if finite, it is unclear how to relate the overall intensity, and thus the number of individuals, N, to the parameters of the log-Gaussian process.

With regard to the specification of the hyperparameters of G_0 (i.e. μ_0, κ_0, λ_0, and ν_0) we note the following: our work has been motivated by data on migratory populations where typically the study encompasses the stopover period and most individuals arrive within the study season. Hence, we defined our prior on phenology to reflect this, as explained in section 2.7. However, the results on the data set of great crested newts, where roughly 40% of the individuals are estimated to have arrival times that are less than one, demonstrate that if the data support it then our model is flexible enough to allow for individuals to arrive before the start of the study. For demonstration purposes, we present a sensitivity analysis for the data set of great crested newts as supplementary material. The analysis suggests that the results, for example the posterior distribution for N and the posterior mean density of arrival times, are robust with respect to the specification of the hyperparameters of G_0 as long as the prior distribution of arrival times does not support arrival that occurs after the end of the study. This is because there are no data available after the end of the study, and hence the posterior will be dominated and completely determined by the prior for that period. As a result, the posterior mean for N will be greater, because N will include individuals that arrived after the end of the study. If this is indeed the prior expectation, as for example suggested by experts, then the results will still be valid. However, in other cases, such as in the case studies of this paper where the expert knowledge suggests that no individuals will arrive after the end of the study, we advise to refrain from specifications of such prior distributions. In our opinion, it is advisable to consider hyperparameters which constain the prior to times that correspond within the study period as this 1) avoids the aforementioned issue of the posterior being dominated by the prior for times when no data are available while 2) does not constrain the posterior to extend to times beyond the study period, or at least before the study commences, if the data suggest so, as demonstrated by our analysis of the great crested newt data set and our sensitivity analysis.

Our approach is generally applicable to data collected on any non-resident wildlife population and our model can be extended in various ways. For example, although the data sets we considered were obtained using only one type of sampling, namely capture, the model can be readily extended for cases when multiple types of sampling are employed, such as capture-resight data. Additionally, the model can be extended for the case of integrated analysis of different (independent) data sets (Besbeas et al., 2002; McCrea et al., 2010; Matechou et al., 2013a; Lyons et al., 2015), to allow for heterogeneity in capture probabilities between individuals (Basu and Ebrahimi, 2001; Rocchetti et al., 2011) and, potentially, to account for misidentification of individuals (McClintock et al., 2014) which is a feature of some non-invasive sampling techniques, such as DNA sampling.

We have chosen to model the departure process using the assumption of a piecewise constant hazard rate which resulted in a modelling framework for (apparent) survival probability similar to that established in the capture-recapture literature. However, more flexible models, for example using continuous kernels as functions of covariates could also be considered. Finally, an interesting extension would be to relate phenology to environmental covariates. One way to address this would be to have the base measure G_0, which tunes the arrival times of each cluster, to be parametrized by these covariates. Alternatively, a different and even more flexible approach would be to consider dependent nonparametric processes (MacEachern, 1999).

We note that there are very few applications of Bayesian nonparametric techniques in population ecology. For example, S. Basu, in an unpublished technical report (Basu, 1998) and Manrique-Vallier (2016) presented a nonparametric Bayesian CR model with heterogeneity in capture probabilities for closed populations based on a Dirichlet process prior while Dorazio et al. (2008) used the same technique to account for heterogeneity in abundance between different sites. However, to our knowledge, the model we presented in this paper is the first Bayesian nonparametric CR model for open populations and we believe that there is great scope for further extension of our work with a considerable range of applications.

Acknowledgements. We acknowledge all the student volunteers who collected the GCN data and Amy Wright, Brett Lewis and Richard Griffiths for collating the data. F. Caron acknowledges the support of the European Commission under the Marie Curie Intra-European Fellowship Programme. We are grateful to Michael Schaub for discussions about the reed warbler data set. We thank the AE and two reviewers for detailed and constructive comments that helped improve the manuscript.

References.

Basu, S. 1998. Capture-recapture and non-parametric Bayes. Technical report, Division of Statistics, Northern Illinois University.

Basu, S. and Ebrahimi, N. 2001. Bayesian capture-recapture methods for error detection and estimation of population size: Heterogeneity and dependence. *Biometrika*, 88(1):269–279.

Bauer, S., Van Dinther, M., Høgda, K.-A., Klaassen, M., and Madsen, J. 2008. The consequences of climate-driven stop-over sites changes on migration schedules and fitness of arctic geese. *Journal of Animal Ecology*, 77(4):654–660.

Besbeas, P., Freeman, S. N., Morgan, B. J. T., and Catchpole, E. A. 2002. Integrating mark–recapture–recovery and census data to estimate animal abundance and demographic parameters. *Biometrics*, 58(3):540–547.

Both, C., Van Turnhout, C. A., Bijlsma, R. G., Siepel, H., Van Strien, A. J., and Foppen, R. P. 2009. Avian population consequences of climate change are most severe for long-distance migrants in seasonal habitats. *Proceedings of the Royal Society of London B: Biological Sciences*, 277(1685):12591266.

Brix, A. 1999. Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability*, 31(4):929–953.

Brix, A. and Diggle, P. J. 2001. Spatiotemporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):823–841.

Charmantier, A. and Gienapp, P. 2014. Climate change and timing of avian breeding and migration: evolutionary versus plastic changes. *Evolutionary Applications*, 7(1):15–28.

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

Daley, D. and Vere-Jones, D. 2008. *An introduction to the theory of point processes*. Springer Verlag.

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

Erni, B., Liechti, F., Underhill, L. G., and Bruderer, B. 2002. Wind and rain govern the intensity of nocturnal bird migration in central europe: a log-linear regression analysis. *Ardea*, 90(1):155–166.

Escobar, M. and West, M. 1995. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90:577–588.

Ferguson, T. 1973. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1:209–230.

Gilbert, S. L., Lindberg, M. S., Hundertmark, K. J., and Person, D. K. 2014. Dead before detection: addressing the effects of left truncation on survival estimation and ecological inference for neonates. *Methods in Ecology and Evolution*, 5(10):992–1001.

Green, P. J. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732.

Ishwaran, H. and James, L. F. 2004. Computational methods for multiplicative intensity models using weighted gamma processes. *Journal of the American Statistical Association*, 99(465):175–190.

Jehle, R., Thiesmeier, B., and Foster, J. 2011. *The Crested Newt: A Dwindling Pond-dweller*. Laurenti Bielefeld.

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

King, R., Morgan, B. J. T., Gimenez, O., and Brooks, S. P. 2009. *Bayesian Analysis for Population Ecology*. Chapman & Hall/CRC, Boca Raton, Florida.

Kingman, J. 1993. *Poisson Processes*, volume 3. Oxford University Press, Great Clarendon Street, Oxford.

Kottas, A. and Sansó, B. 2007. Bayesian mixture modeling for spatial Poisson process intensities, with applications to extreme value analysis. *Journal of Statistical Planning and Inference*, 137(10):3151–3163.

Kovács, S., Fehérvári, P., Nagy, K., Harnos, A., and Csörgő, T. 2012. Changes in migration phenology and biometrical traits of reed, marsh and sedge warblers. *Open Life Sciences*, 7(1):115–125.

Kuo, L. and Ghosh, S. K. 1997. Bayesian nonparametric inference for nonhomogeneous Poisson processes. Technical report, University of Connecticut, Department of Statistics.

Lebreton, J.-D., Burnham, K. P., Clobert, J., and Anderson, D. R. 1992. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological monographs*, 62(1):67–118.

Lewis, B. 2012. *An evaluation of mitigation actions for great crested newts at development sites*. PhD thesis, Durrell Institute of Conservation and Ecology, School of Anthropology & Conservation, University of Kent.

Lo, A. and Weng, C.-S. 1989. On a class of Bayesian nonparametric estimates: II.Hazard rate estimates. *Annals of the Institute of Statistical Mathematics*, 41(2):227–245.

Lyons, J. E., Kendall, W. L., Royle, J. A., Converse, S. J., Andres, B. A., and Buchanan, J. B. 2015. Population size and stopover duration estimation using mark–resight data and Bayesian analysis of a superpopulation model. *Biometrics*.

MacEachern, S. N. 1994. Estimating normal means with a conjugate style Dirichlet process prior. *Communications in Statistics-Simulation and Computation*, 23(3):727–741.

MacEachern, S. N. 1999. Dependent nonparametric processes. In *ASA proceedings of the section on Bayesian statistical science*, pages 50–55. Alexandria, Virginia. Virginia: American Statistical Association; 1999.

Manrique-Vallier, D. 2016. Bayesian population size estimation using dirichlet process mixtures. *Biometrics*.

Matechou, E., Morgan, B. J. T., Pledger, S., Collazo, J. A., and Lyons, J. E. 2013a. Integrated analy-

sis of capture–recapture–resighting data and counts of unmarked birds at stop-over sites. *Journal of Agricultural, Biological, and Environmental Statistics*, 18(1):120–135.

Matechou, E., Pledger, S., Efford, M., Morgan, B. J. T., and Thomson, D. L. 2013b. Estimating age-specific survival when age is unknown: open population capture–recapture models with age structure and heterogeneity. *Methods in Ecology and Evolution*, 4(7):654–664.

McClintock, B. T., Bailey, L. L., Dreher, B. P., Link, W. A., et al. 2014. Probit models for capture–recapture data subject to imperfect detection, individual heterogeneity and misidentification. *The Annals of Applied Statistics*, 8(4):2461–2484.

McCrea, R. S., Morgan, B. J., Gimenez, O., Besbeas, P., Lebreton, J.-D., and Bregnballe, T. 2010. Multi-site integrated population modelling. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(4):539–561.

Møller, J. 2003. Shot noise Cox processes. *Advances in Applied Probability*, 35:614–640.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. 1998. Log Gaussian Cox processes. *Scandinavian journal of statistics*, 25(3):451–482.

Møller, J. and Waagepetersen, R. P. 2004. *Statistical inference and simulation for spatial point processes*. CRC Press.

Neal, R. 2000. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9:249–265.

Nieto-Barajas, L. and Walker, S. G. 2004. Bayesian nonparametric survival analysis via Lévy driven Markov processes. *Statistica Sinica*, 14(4):1127–1146.

Pitman, J. 1996. Some developments of the Blackwell-Macqueen urn scheme. *Lecture Notes-Monograph Series*, pages 245–267.

Pledger, S., Efford, M., Pollock, K. H., Collazo, J. A., and Lyons, J. E. 2009. Stopover duration analysis with departure probability dependent on unknown time since arrival. *Environmental and Ecological Statistics (Edited by D.L.Thomson, E.G.Cooch and M.J. Conroy)*, 3:349–363.

Plummer, M., Best, N., Cowles, K., and Vines, K. 2006. Coda: Convergence diagnosis and output analysis for mcmc. *R News*, 6(1):7–11.

Pradel, R. 1996. Utilization of capture-mark-recapture for the study of recruitment and population growth rate. *Biometrics*, 52:703–709.

R Core Team 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rocchetti, I., Bunge, J., and Böhning, D. 2011. Population size estimation based upon ratios of recapture probabilities. *The Annals of Applied Statistics*, pages 1512–1533.

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

Royle, J. A., Karanth, K. U., Gopalaswamy, A. M., and Kumar, N. S. 2009. Bayesian inference in camera trapping studies for a class of spatial capture-recapture models. *Ecology*, 90(11):3233–3244.

Royle, J. A. and Young, K. V. 2008. A hierarchical model for spatial capture-recapture data. *Ecology*, 89(8):2281–2289.

Schaub, M., Liechti, F., and Jenni, L. 2004. Departure of migrating european robins, erithacus rubecula, from a stopover site in relation to wind and rain. *Animal Behaviour*, 67(2):229–237.

Schaub, M., Pradel, R., Jenni, L., and Lebreton, J.-D. 2001. Migrating birds stop over longer than usually thought: an improved capture-recapture analysis. *Ecology*, 82(3):852–859.

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

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

Seebacher, F. and Post, E. 2015. Climate change impacts on animal migration. *Climate Change Responses*, 2(1):1.

Sethuraman, J. 1994. A constructive definition of Dirichlet priors. *Statistica Sinica*, 4:639–650.

Sullivan, A. R., Flaspohler, D. J., Froese, R. E., and Ford, D. 2015. Climate variability and the timing of

spring raptor migration in eastern north america. *Journal of Avian Biology*.
Taddy, M. A. and Kottas, A. 2012. Mixture modeling for marked Poisson processes. *Bayesian Analysis*, 7(2):335–362.
Van Buskirk, J., Mulvihill, R. S., and Leberman, R. C. 2009. Variable shifts in spring and autumn migration phenology in north american songbirds associated with climate change. *Global Change Biology*, 15(3):760–771.
Wolpert, R. L. and Ickstadt, K. 1998. Poisson/gamma random field models for spatial statistics. *Biometrika*, 85(2):251–267.

5. Appendix A: details on the gamma process. A draw from a gamma process is an almost surely discrete measure, and takes the following form:

$$G = \omega \sum_{j=1}^{\infty} \pi_j \delta_{(\mu_j^\star, \sigma_j^{\star 2})}, \tag{5.1}$$

where δ_a is the Dirac delta measure at a. Combining Eq. (5.1) and (2.1) we obtain the following infinite mixture of Gaussian form for the unknown intensity ν

$$\nu(\zeta|G) = \omega \sum_{j=1}^{\infty} \pi_j \mathcal{N}(\zeta; \mu_j^\star, \sigma_j^{\star 2}). \tag{5.2}$$

The $(\pi_j)_{j=1,2,\ldots}$ are positive weights which sum to one and follow a stick-breaking process (Sethuraman, 1994) with $\pi_j = \theta_j \prod_{\ell=1}^{j-1}(1-\theta_\ell)$ where $\theta_j \sim \text{Beta}(1,\alpha)$. The positive scaling variable ω has distribution $\omega \sim \text{Gamma}(\alpha,\tau)$ while the mixture means and variances $(\mu_j^\star, \sigma_j^{\star 2})$, are i.i.d. from G_0.

The above construction can be further simplified by the introduction of a suitable set of latent variables and the use of the remarkable conjugacy properties of the gamma process. Given G, the arrival times $(\zeta_i)_{i=1,\ldots,N}$ are drawn from a Poisson process with intensity $\nu(\zeta|G)$, or

$$N|G \sim \text{Poisson}(\omega)$$

and for $i = 1, \ldots, N$

$$\zeta_i|G \overset{\text{i.i.d.}}{\sim} \frac{\nu(\zeta|G)}{\omega}. \tag{5.3}$$

As the intensity ν takes the mixture form (2.1), (5.3) can be alternatively represented in the following hierarchical form, for $i = 1, \ldots, N$,

$$(\tilde{\mu}_i, \tilde{\sigma}_i^2)|G \sim \overline{G} \tag{5.4}$$

$$\zeta_i|(\tilde{\mu}_i, \tilde{\sigma}_i^2) \sim \mathcal{N}(\tilde{\mu}_i, \tilde{\sigma}_i). \tag{5.5}$$

where $\overline{G} = G/\omega$ and the $(\tilde{\mu}_i, \tilde{\sigma}_i^2), i = 1, \ldots, N$ are latent variables indicating the mean and variance of the Gaussian component from which ζ_i originated. As G is almost surely discrete, the latent variables may have duplicate values. We write $(\mu_j, \sigma_j^2)_{j=1,\ldots,J}$ the set of unique values in $(\tilde{\mu}_i, \tilde{\sigma}_i^2)_{i=1,\ldots,N}$, and $\Pi_N = \{A_1, \ldots, A_J\}$ the partition (or clustering) of the N individuals, such that individuals k and ℓ are in the same cluster iff $(\tilde{\mu}_k, \tilde{\sigma}_k^2) = (\tilde{\mu}_\ell, \tilde{\sigma}_\ell^2)$. $J \leq N$ is the number of different non-empty clusters.

As $\overline{G}$ is obtained by normalization of a gamma process, it is distributed from a Dirichlet process (Ferguson, 1973; Kingman, 1993); using the conjugacy properties of the Dirichlet process (Kingman, 1993; Pitman, 1996), it is actually possible to analytically integrate out $\overline{G}$. The associated marginal distribution over the partition Π_N of the N arrival times, is given by

$$\Pr(\Pi_N = \{A_1, \ldots, A_J\} | \alpha, N) = \frac{\Gamma(\alpha)}{\Gamma(\alpha + N)} \alpha^J \prod_{j=1}^{J} \Gamma(n_j) \tag{5.6}$$

where $n_j = \text{card}(A_j)$, $j = 1, \ldots, J$ is the size of cluster j. The generative process for such partition is known as the Chinese restaurant process (CRP):

$$\Pr(\text{individual } N+1 \text{ joins an existing cluster } j | \Pi_N) = \frac{n_j}{\alpha + N}, \ j = 1, \ldots, J$$

$$\Pr(\text{individual } N+1 \text{ joins a new cluster} | \Pi_N) = \frac{\alpha}{\alpha + N} \tag{5.7}$$

This marginalization is important in practice for MCMC inference, as it allows us to perform inference with a set of parametric parameters, although the model actually involves an infinite-dimensional parameter.

6. Appendix B: Overall hierarchical model. Let J be the number of clusters in Π_N. Let $c_i \in \{1, \ldots, J\}$ indicate the index of the cluster to which individual i belongs, i.e. $i \in A_{c_i}$. The overall model can be described as

$$\alpha \sim \text{Gamma}(a_\alpha, b_\alpha) \quad \text{[Tunes the nb of clusters and overall intensity]} \tag{6.1}$$

$$\tau \sim \text{Gamma}(a_\tau, b_\tau) \quad \text{[Tunes the overall intensity]} \tag{6.2}$$

$$\beta \sim \mathcal{N}(0_q, I_q) \quad \text{[Coefficients for capture]} \tag{6.3}$$

$$\gamma \sim \mathcal{N}(0_{\tilde{q}}, I_{\tilde{q}}) \quad \text{[Coefficients for departure]} \tag{6.4}$$

$$\omega | \alpha, \tau \sim \text{Gamma}(\alpha, \tau) \quad \text{[Overall intensity of the arrival process]} \tag{6.5}$$

$$N | \omega \sim \text{Poisson}(\omega) \quad \text{[Overall size of the population]} \tag{6.6}$$

$$\Pi_N | N, \alpha \sim (5.6) \quad \text{[Partition of the individuals]} \tag{6.7}$$

for $j = 1, 2, \ldots, J$

$$(\mu_j, \sigma_j^2) \sim G_0 \quad \text{[Means and variances of the clusters]} \tag{6.8}$$

and for $i = 1, 2, \ldots, N$

$$\zeta_i | c_i, \mu_{c_i}, \sigma^2_{c_i} \sim \mathcal{N}(\mu_{c_i}, \sigma^2_{c_i}) \quad \text{[Arrival times]} \tag{6.9}$$

$$d_i | \zeta_i, \gamma \sim \Pr(d_i | \zeta_i, \gamma) \quad \text{[Departure indices]} \tag{6.10}$$

$$\mathrm{H}_i \, | \zeta_i, d_i, \beta \sim \Pr(\mathrm{H}_i \, | \zeta_i, d_i, \beta) \quad \text{[Capture histories]} \tag{6.11}$$

7. Appendix C: Posterior inference. The vector of unknown parameters consists of:

$$\alpha, \tau, \omega, N, \Pi_N, (\mu_{1:J}, \sigma^2_{1:J}), \zeta_{1:N}, d_{1:N}, \beta \text{ and } \gamma.$$

Our objective is to approximate the posterior distribution

$$P(\alpha, \tau, \omega, N, \Pi_N, \zeta_{1:N}, d_{1:N}, \mu_{1:J}, \sigma^2_{1:J}, \beta, \gamma | \mathcal{D}).$$

This distribution is not analytically tractable, and we resort to an MCMC algorithm to provide samples asymptotically distributed from it. Note that the number of clusters J is not set in advance, and may change at each iteration of the algorithm.

We use indices $1, \ldots, D$ to indicate individuals observed, and $D+1, \ldots, N$ for individuals unobserved. The MCMC algorithm iterates as follows:

1. Jointly update (α, ω):
   (a) First we sample α given all variables except ω (see details below):
   $$\alpha | \text{rest except } \omega \sim \text{Gamma}\left(J + a_\alpha, \log\left(1 + \frac{1}{\tau}\right) + b_\alpha\right)$$
   (b) Then we sample ω given the rest:
   $$\omega | \text{rest} \sim \text{Gamma}(\alpha + N, \tau + 1)$$
2. Update τ:
   $$\tau | \alpha, \omega \sim \text{Gamma}(a_\tau + \alpha, b_\tau + \omega).$$
3. For $i = 1, \ldots, N$, update c_i:
   This update is the MCMC update for conjugate models in Dirichlet process mixtures (MacEachern, 1994; Escobar and West, 1995), see e.g. Neal (2000, Algorithm 3).
   For a subset of the arrival times $S \subseteq \{1, \ldots, N\}$ let $\zeta_S = \{\zeta_i | i \in S\}$ and
   $$f(\zeta_S) = \int_{-\infty}^{\infty} \int_0^{\infty} \prod_{i \in S} \mathcal{N}(\zeta_i; \mu, \sigma^2) G_0(d\mu, d\sigma^2)$$

As the normal inverse gamma distribution G_0 is a conjugate prior for a normal likelihood, $f(\zeta_S)$ can be evaluated analytically. Then, let $\mathcal{I}$ be the set of different indices of c_j, $j \neq i$, and for $c \in \mathcal{I}$ let $S_{c,-i}$ be the set of individuals $j \neq i$ in cluster c and $n_{c,-i} = \text{card}(S_{c,-i})$ the number of individuals $j \neq i$ in cluster c. Then individual i will join an existing cluster $c \in \mathcal{I}$ with probability

$$\Pr(c_i = c|\text{rest}) \propto n_{c,-i} \frac{f(\zeta_{S_{c,-i} \cup \{i\}})}{f(\zeta_{S_{c,-i}})}$$

or be allocated to a new cluster with probability

$$\Pr(c_i = \text{new}|\text{rest}) \propto \alpha f(\zeta_{\{i\}}).$$

Note that at the end of this step, we obtain an updated partition Π_N with a potentially different number of clusters J.

4. For $j = 1, \ldots, J$ update (μ_j, σ_j^2):
These are updated for cluster j, $j = 1, \ldots, J$, as:

$$\mu_j|\sigma_j^2 \sim \mathcal{N}\left(m_j, \sigma_j^2/\kappa_j\right) \tag{7.1}$$

$$1/\sigma_j^2 \sim \text{Gamma}(\nu_j, \lambda_j) \tag{7.2}$$

where

$$\begin{aligned}
\nu_j &= \nu_0 + n_j/2 \\
\kappa_j &= \kappa_0 + n_j \\
m_j &= \frac{\kappa_0 m_0 + n_j \overline{\zeta_j}}{\kappa_0 + n_j} \\
\lambda_j &= \lambda_0 + \frac{1}{2} \sum_{i=1}^{n_j} (\zeta_i - \overline{\zeta_j})^2 + \frac{n_j \kappa_0}{n_j + \kappa_0} \frac{(\overline{\zeta_j} - m_0)^2}{2}
\end{aligned}$$

with $\overline{\zeta_j} = \frac{1}{n_j} \sum_{i|c_i=j} \zeta_i$ and $n_j = \text{card}(\{i|c_i = j\})$.

5. For $i = 1, \ldots, N$, update ζ_i:
We use a random walk Metropolis-Hastings step where we propose to update ζ_i to $\zeta_i' = \zeta_i + \text{Gaussian noise}$. The acceptance probability is:

$$\min\left(1, \frac{\Pr(\zeta_i'; \mu_{c_i}, \sigma_{c_i}^2) \Pr(\mathrm{H}_i\,|\zeta_i', d_i, \beta) \Pr(d_i|\zeta_i', \gamma)}{\Pr(\zeta_i; \mu_{c_i}, \sigma_{c_i}^2) \Pr(\mathrm{H}_i\,|\zeta_i, d_i, \beta) \Pr(d_i|\zeta_i, \gamma)}\right)$$

6. For $i = 1, \ldots, N$, update d_i.
We use a Metropolis-Hastings step where we propose either $d_i' = d_i + 1$ or $d_i' = d_i - 1$ with equal probability, unless $d_i = 0$ when $d_i' = 1$ or $d_i = K$ when $d_i' = K - 1$ with probability 1. Hence, $q(d_i'|d_i) = 1/2$ for $d_i = 1, \ldots, K-1$, $q(d_i' = 1|d_i = 0) = q(d_i' = K-1|d_i = K) = 1$ and 0 otherwise while the same holds for $q(d_i|d_i')$. The acceptance probability is:
$$\min\left(1, \frac{\Pr(\mathrm{H}_i\,|\zeta_i, d_i', \beta)\Pr(d_i'|\zeta_i, \gamma)q(d_i'|d_i)}{\Pr(\mathrm{H}_i\,|\zeta_i, d_i, \beta)\Pr(d_i|\zeta_i, \gamma)q(d_i|d_i')}\right)$$

7. Update β and γ:
These are coefficients of logistic regression models so their update is performed using a Metropolis-Hastings algorithm, described for example in Chapter 8 of King et al. (2009).
We note here that for the Metropolis-Hastings steps the proposal variances were chosen after tuning i.e. running a small number of trial runs and visually inspecting the resulting trace plots for good mixing of the chain.

8. Update $N, c_{D+1:N}, \zeta_{D+1:N}, d_{D+1:N}$:
The colouring theorem for marked Poisson processes implies that, conditional on G, the set of points $\{\zeta_{D+1:N}, d_{D+1:N}\}$ is independent of $\{\zeta_{1:D}, d_{1:D}\}$ and distributed from a non-homogeneous Poisson process with intensity
$$\nu_0(\zeta, d|G) = \nu(\zeta|G)\Pr(d|\zeta, \gamma)\Pr(\mathrm{H} = (0, \ldots, 0)|\zeta).$$
As the normalized measure $\overline{G}$ is marginalized out, some dependency is retained through the cluster variables $c_{1:D}$. We sample from the conditional distribution of $(N, c_{D+1:N}, \zeta_{D+1:N}, d_{D+1:N})$ given the rest by rejection as follows.
First, sample $N_0 \sim \mathrm{Poisson}(\omega)$. For $i = 1, \ldots, N_0$, sample the latent cluster variables from the Chinese restaurant process
$$c_i^*|c_{1:D}, c_{1:i-1}^* \sim (5.7)$$
and whenever c_i^* takes a new value, sample the new cluster location from G_0.
For $i = 1, \ldots, N_0$, sample
$$\begin{aligned}\zeta_i^*|c_i^*, \mu, \sigma^2 &\sim \mathcal{N}(\mu_{c_i^*}, \sigma_{c_i^*}^2)\\ d_i^*|\zeta_i^*, \gamma &\sim \Pr(d_i^*|\zeta_i^*, \gamma)\\ \mathrm{H}_i^*\,|d_i^*, \zeta_i^*, \beta &\sim \Pr(\mathrm{H}_i^*\,|d_i^*, \zeta_i^*, \beta)\end{aligned}$$
We only keep the $\widetilde{N}_0 \leq N_0$ individuals i for which $\mathrm{H}_i^* = (0, 0, \ldots, 0)$, set $N = D + \widetilde{N}_0$ and relabel them from $D+1$ to N to obtain updated values $(c_{D+1:N}, \zeta_{D+1:N}, d_{D+1:N})$.

Details of Step 1(a). We have

$$\begin{aligned}\Pr(N|\alpha) &= \int_0^\infty \Pr(N,\omega|\alpha)d\omega \\ &= \int_0^\infty \Pr(N|\omega,\alpha)p(\omega|\alpha)d\omega \\ &= \frac{\Gamma(N+\alpha)\tau^\alpha}{N!\Gamma(\alpha)(1+\tau)^{N+\alpha}} \\ &= \frac{\Gamma(N+\alpha)}{N!\Gamma(\alpha)}\exp\left\{-\alpha\log\left(1+\frac{1}{\tau}\right)-N\log(1+\tau)\right\}\end{aligned}$$

which gives, together with Eq. (5.6) and the gamma prior on α

$$\begin{aligned}\Pr(\alpha|\text{rest except }\omega) &\propto \Pr(\Pi_N|\alpha,N)\Pr(N|\alpha)p(\alpha) \\ &\propto \alpha^{J+a_\alpha-1}\exp\left[-\alpha\left\{\log\left(1+\frac{1}{\tau}\right)+b_\alpha\right\}\right].\end{aligned}$$

E-mail: E.Matechou@kent.ac.uk

School of Mathematics, Statistics & Actuarial Science,
University of Kent, Canterbury, UK,
E-mail: E.Matechou@kent.ac.uk

E-mail: Caron@stats.ox.ac.uk

Department of Statistics,
University of Oxford, Oxford, UK
E-mail: Caron@stats.ox.ac.uk