Environ Ecol Stat (2016) 23:531–547
DOI 10.1007/s10651-016-0352-0

Bayesian analysis of Jolly-Seber type models

Incorporating heterogeneity in arrival and departure

Eleni Matechou[1] · Geoff K. Nicholls[2] · Byron J. T. Morgan[1] · Jaime A. Collazo[3] · James E. Lyons[4]

Received: 18 March 2015 / Revised: 30 May 2016 / Published online: 4 August 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract We propose the use of finite mixtures of continuous distributions in modelling the process by which new individuals, that arrive in groups, become part of a wildlife population. We demonstrate this approach using a data set of migrating semipalmated sandpipers (*Calidris pussila*) for which we extend existing stopover models to allow for individuals to have different behaviour in terms of their stopover duration at the site. We demonstrate the use of reversible jump MCMC methods to derive posterior distributions for the model parameters and the models, simultaneously. The algorithm moves between models with different numbers of arrival groups as well as between models with different numbers of behavioural groups. The approach is shown to provide new ecological insights about the stopover behaviour of semipalmated sandpipers but is generally applicable to any population in which animals arrive in groups and potentially exhibit heterogeneity in terms of one or more other processes.

Handling Editor: Bryan F. J. Manly.

Electronic supplementary material The online version of this article (doi:10.1007/s10651-016-0352-0) contains supplementary material, which is available to authorized users.

✉ Eleni Matechou
e.matechou@kent.ac.uk

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

2 Department of Statistics, Oxford University, Oxford, UK

3 U.S. Geological Survey, North Carolina Cooperative Fish and Wildlife Research Unit, Department of Applied Ecology, North Carolina State University, Raleigh, USA

4 U.S. Fish and Wildlife Service, Division of Migratory Bird Management, Patuxent Wildlife Research Center, Laurel, MD 20708, USA

Keywords Capture-recapture-resight data sets · Integrated modelling · Mixture models · Reversible jump · Semipalmated sandpipers · Stopover data

1 Introduction

Capture-recapture (CR) data, arising when individuals are captured, individually marked, and followed over time, are often collected from wildlife populations. In some cases, more than one type of sampling is used, as in capture-recapture-resight (CRR) data when individuals can be detected without necessarily being caught or in capture-recovery data when individuals can be detected dead. Different sampling schemes can also result in more than one data set being collected from the same population, and these are often analysed using an integrated approach (Besbeas et al. 2002).

Population ecology models are employed to analyse CR, CRR etc. data in order to estimate, among other things, the size of the population and the probabilities of survival of the individuals. These models need to account for the sampling scheme, imperfect detection and often also for potential heterogeneity between individuals. This can arise either in their characteristics, for example survival probabilities, or in their behaviour, that could for example affect their detection probability.

Population ecology models are referred to as Jolly-Seber (JS) type (Jolly 1965; Seber 1965), if they model the process by which new individuals enter the population. An example of a JS type model is the Schwarz and Arnason (1996) model, which uses the idea of a super-population of animals, N, and the entry parameters, $\beta_{b-1},\ b = 1, \ldots, K$ to denote the proportion of N that were new arrivals on sampling occasion b, where K is the number of samples. For a discussion of alternative JS type model formulations see Sect. 8.2.3 in McCrea and Morgan (2014).

Individuals enter the population either through birth or immigration and they often do so in groups. For example, migrating birds arrive at stopover sites in flocks rather than individually while juveniles of a species can emerge in a synchronous manner. If the number of arrival or emergence groups is known, then finite mixture models of continuous distributions, such as the normal, can be used to model the process by which new individuals enter the population. See for example Matechou et al. (2014) who modelled the emergence of butterfly broods using a mixture of two normal distributions.

However in many cases the number of arrival groups, and hence the number of mixture components, is unknown. Model selection criteria, such as the Akaike information criterion (Akaike 1973, AIC), have doubtful validity for selecting between models with different mixture components (see McLachlan and Peel 2000, Chap. 6). Pledger et al. (2010) refer to a comment by Burnham and Anderson (2002) who suggest that the parameter estimates have to be in the interior of the parameter space for AIC to be valid. Cubaynes et al. (2012) report relatively low success rates of AIC, the Bayesian information criterion (Schwarz 1978, BIC) and the integrated classification criterion (ICL–BIC), which is similar to BIC but has an additional penalty for fuzzy clustering (Biernacki et al. 2000), in selecting the true number of mixture components in CR data. Finally, parameter estimates can be sensitive to model choice

(Pledger 2000) and choosing a single model for inference can be undesirable as well as difficult.

Arnold et al. (2010) demonstrated the use of reversible jump (Green 1995, RJ) MCMC in the case of finite mixture models that are used to account for heterogeneity in capture probabilities in closed populations. They referred to RJMCMC as a useful means of selecting between models with different numbers of mixture components, or obtaining model-averaged estimates of parameters. RJMCMC has also been used in the capture-recapture literature (Brooks et al. 2000; King and Brooks 2008; King et al. 2010, for example) as a method for selecting model covariates, assessing whether model parameters are constant over time or time-varying or to compare models which allow for heterogeneity between individuals using random effects to models which assume a homogeneous population.

In this paper we demonstrate the use of finite mixture models to describe the arrival pattern of migrating semipalmated sandpipers (*Calidris pussila*) at a stopover site in terms of mixtures of continuous distributions, specifically the normal distribution. This approach provides a biologically meaningful interpretation of the results in which each mixture component is treated as a flock, so that flocks can be compared in terms of their relative sizes and mean arrival times. We use RJMCMC to obtain the posterior distribution of the number of arrival groups and a model-averaged estimate of the arrival pattern.

The data set of semipalmated sandpipers was first analysed in Matechou et al. (2013a) (M13) who extended the stopover model of Pledger et al. (2009) by proposing integrated models for stopover data on birds that are marked, and therefore individually identifiable, together with raw count data of unmarked birds. They modelled the probability that an individual present at the stopover site will remain until the next sampling occasion, termed retention probability, as a function of calendar time and of the unknown time the individual has already spent at the site, which they referred to as its "age". These stopover models provide estimates of the population size and indirect estimates of the total stopover duration. Stopover sites provide an essential opportunity for migrating birds to break their journey, rest and refuel. It is important to assess the significance of a site, an attribute which is based on the number of migrants that use it and the duration of their stopover, as this can aid in formulating conservation strategies aimed at non-breeding habitat for migrant shorebirds (Brown et al. 2001) and in measuring the effects of management treatments (Nichols and Williams 2006; Lyons et al. 2008).

However, in M13 all birds are assumed to behave independently and identically in terms of their stopover duration, a feature which is known to be untrue for many migratory species (Alerstam and Lindström 1990; Lyons and Haig 1995; Cristol et al. 1999; Dinsmore and Collazo 2003; Rubolini et al. 2004; Bishop et al. 2006). In this paper we also use finite mixtures to allow for different behavioural groups, defined by their retention probability and hence stopover duration at the site. Our results agree with life history strategies (e.g., mating system) that allow for differential migration strategies among sexes to maximize fitness (Rubolini et al. 2004) and show that there are at least two behavioural groups of semipalmated sandpipers.

Hence, the work in this paper demonstrates the use of RJMCMC that moves between finite mixture models with different numbers of homogeneous groups in two directions: arrival groups and behavioural groups. By using RJMCMC we are able to quantify the uncertainty arising from the need to estimate the number of mixture components in each direction instead of relying on model selection criteria to choose the "best" model. Additionally, the posterior distributions of model parameters, or functions of them, can be naturally averaged across the different models, if appropriate and desirable.

In Sect. 2 we give a brief introduction to finite mixture models and the RJMCMC algorithm. We present the data set of semipalmated sandpipers and the results in Sect. 3. The details of the RJMCMC algorithm specific to the application and convergence diagnostic checks are given as Supplementary Material.

We verified our formulae and code by comparing our results to those obtained from a very simple but reliable and independently coded rejection algorithm, suitable for simple data sets only. In general terms, this works as follows: for very small discrete data sets Y, we can simulate the posterior using a very simple rejection algorithm which simulates parameters θ from the prior, and then simulates synthetic data $Y'|\theta$ according to the observation model, and finally accepts θ as an independent sample from the posterior if $Y' = Y$, where Y is the real data. This exact algorithm is actually the motivating algorithm for ABC (Tavaré et al. 1997). We also fitted synthetic data of similar size to the real data set and checked convergence of our algorithm to known parameter values. We present the results of a simulation study as Supplementary Material.

2 Mixture models and RJMCMC

The data are represented in $\mathbf{X}$ with $\mathbf{X}_i$ the ith data vector. In generic mixture models, the model parameters are:

- G, the number of mixture components,
- $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_G)$, $\sum_{g=1}^{G} \pi_g = 1$, the mixing proportions,
- $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_G)$ with $\boldsymbol{\eta}_g$ the collection of parameters of the gth mixture component, and
- $\boldsymbol{\psi}$, a collection of parameters that are not part of the mixture components.

We write $\boldsymbol{\theta} = (G, \boldsymbol{\pi}, \boldsymbol{\eta}, \boldsymbol{\psi})$. We seek an expression for the posterior distribution,

$$\mathcal{P}(\boldsymbol{\theta}|\mathbf{X}) \propto \mathcal{P}(\mathbf{X}|\boldsymbol{\theta})\mathcal{P}(\boldsymbol{\theta})$$

of the parameters in $\boldsymbol{\theta}$, allowing the number of mixture components, G, and hence the number of parameters in $\boldsymbol{\theta}$ to be estimated, where

$$\mathcal{P}(\mathbf{X}|\boldsymbol{\theta}) = \prod_i \sum_{g=1}^{G} \pi_g \mathcal{P}(\mathbf{X}_i|\boldsymbol{\eta}_g, \boldsymbol{\psi})$$

and $\mathcal{P}(\boldsymbol{\theta})$ is the joint prior of the parameters.

We summarise $\mathcal{P}(\boldsymbol{\theta}|\mathbf{X})$ using a RJMCMC algorithm. This has two update types: one for updating parameters within models, $\boldsymbol{\pi}$, $\boldsymbol{\eta}$, $\boldsymbol{\psi}$, and one for updating the number of mixture components, G.

We update within-model parameters $\boldsymbol{\eta}$ and $\boldsymbol{\psi}$ using a standard single-update Metropolis-Hastings random walk, described for example in King et al(2010, Sect. 5.3.2). Mixing proportions, $\boldsymbol{\pi}$, are updated as follows: two groups are chosen at random, say g_1 and g_2, ϵ is defined as $\epsilon = \gamma(\pi_{g_1} + \pi_{g_2})$, where $\gamma \in (0, 1)$ is fixed and chosen during tuning, x is drawn from Unif(-ϵ, ϵ) and π'_{g_1} and π'_{g_2} are calculated by $\pi'_{g_1} = \pi_{g_1} + x$ and $\pi'_{g_2} = \pi_{g_2} - x$. If $\pi'_{g_1}, \pi'_{g_2} \geq 0$ and $\pi'_{g_1} \leq (\pi_{g_1} + \pi_{g_2})$ the standard Metropolis-Hastings acceptance probability is calculated.

The number of mixture components, G, is updated using a RJMCMC move. The proposal transition probability to a model with G' mixture components and $\boldsymbol{\theta}'$ parameters from a model with G components and $\boldsymbol{\theta}$ parameters is denoted by $P_G(G'|G)$.

Suppose that the proposed move is to a model with $G' = G + 1$ groups. We allocate mass to this newly formed group by removing some mass from an existing group. Specifically, the proposed proportion of individuals in this new group, π'_{G+1}, is generated by choosing one of the existing G groups at random, say group g, with probability $1/G$, drawing x from Unif(0, π_g), setting π'_{G+1} equal to x and π'_g equal to $\pi_g - x$.

The parameters for this proposed group, $\boldsymbol{\eta}'_{G+1}$, are generated from their corresponding prior, $\mathcal{P}(\boldsymbol{\eta}'_{G+1})$.

Suppose that the proposed move is to a model with $G' = G - 1$ groups. We choose g_1 from Unif$\{1, \ldots, G\}$ and g_2 from Unif$\{1, \ldots, g_1 - 1, g_1 + 1, \ldots, G\}$. We remove group g_1 and allocate its mass to group g_2.

The acceptance probability for a model with $G' = G + 1$ groups is given by

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min\left(1, \frac{\mathcal{P}(\boldsymbol{\theta}'|\mathbf{X}) P_G(G|G+1) \frac{1}{(G+1)} \frac{1}{G}}{\mathcal{P}(\boldsymbol{\theta}|\mathbf{X}) P_G(G+1|G) \frac{1}{G} \frac{1}{\pi_{g_1}} \mathcal{P}(\boldsymbol{\eta}'_{G+1})}\right). \tag{1}$$

The Jacobian term (see King et al. 2010, p. 165) required in forming equation (1) is equal to 1 because G' and π'_{g_1}, π'_{G+1} are, respectively, linear functions of G and π_{g_1} and $\boldsymbol{\eta}'_{G+1}$ are generated from their prior.

The reverse move, to a model with $G - 1$ groups, is fully defined given the above and is presented in detail for the example considered in this paper as Supplementary Material.

Arnold et al. (2010) provide a detailed description of RJMCMC for closed population models that allow for heterogeneity in capture probability and we provide R (R Core Team 2015) code and details of the algorithm for analysing a data set from a closed population of cottontail rabbits, also presented by Arnold et al. (2010) as Supplementary Material. For the application we present in this paper, our population is instead open and exhibits heterogeneity in both arrival and departure. We give details of the algorithm in this case as Supplementary material and we make R code available on request from the first author.

We check convergence of the algorithm by running three chains with random starting values for the parameters and using convergence diagnostic checks incorporated

in the R package *coda* (Plummer et al. 2006). These are presented as Supplementary Material. Accurate implementation of RJMCMC is non-trivial here due in part to the need to include normalizing constants, which cancel in fixed dimension Metropolis Hastings MCMC but not in RJMCMC. We checked the accuracy of our code by making an independent implementation using a very simple second approach based on rejection. This second approach is relatively inefficient and would not scale to data sets of practical interest, but does allow us to make high precision checks on small data sets. The posterior distribution was simulated by the two methods and the results were in excellent agreement.

3 Application

3.1 Data and parameters

The stopover site is formed by the wetlands at the Tom Yawkey Wildlife Center in South Carolina where the study, which spans $T = 38$ days, took place in spring of 2001. Samples are collected on $K = 29$ of these days and there are nine null occasions when no sampling takes place. There are two types of sampling occasions: on capture occasions, of which there are 11, birds can be caught using mist nets and uniquely marked before being released; on resight occasions, of which there are 18, marked birds can be detected and an imperfect count of unmarked birds is obtained. These raw counts of unmarked birds form a vector $\mathbf{y}$ of length T with 18 non-missing entries corresponding to resighting occasions.

Each of the birds that visited the site during the study has its own capture-recapture-resight history (CRRH) and we let H denote the number of distinct observed CRRHs of the D birds that were marked. For CRRH $\mathbf{x}_h = (x_{h1}, \ldots, x_{hT})$, shared by n_h birds, with $x_{ht} \in \{0, 1, 2\}$, 2, 1, and 0 signifying that the n_h individuals were resighted, caught or missed, respectively, on occasion t. Any bird that was never caught has the trivial history $\mathbf{0}$. All CRRHs have nine missing entries. The H CRRHs are summarised in matrix $\mathbf{X}$ of dimension $H \times T$ and their frequencies are recorded in vector $\mathbf{n}$.

The $\mathbf{y}$–data, formed by the raw counts, and the $\mathbf{X}$–data, formed by the H unique CRRHs together with their frequencies in $\mathbf{n}$, are the two data sets to be analysed using the M13 proposed integrated model which has two parts: one that builds on the Pledger et al. (2009) model for the $\mathbf{X}$–data and a binomial model for the $\mathbf{y}$–data.

The model parameters are:

- N: super-population size. The total number of birds that became available for capture-resight during the study without necessarily being detected.
- M: number of arrival groups.
- w_m, μ_m and σ_m, $m = 1, \ldots, M$: respectively, population fractions, mean arrival times and standard deviations of arrival times of the M arrival groups, $\sum_{m=1}^{M} w_m = 1$. The population fraction that arrived between occasions $b - 1$ and b is the entry parameter β_{b-1}. In terms of the mixture components,

$$\beta_{b-1} = \sum_{m=1}^{M} w_m \left\{ F_m(b) - F_m(b-1) \right\}, \; b = 2, \ldots, T-1,$$

where $F_m(b) = P(X \leq b)$ when $X \sim N(\mu_m, \sigma_m^2)$. The first and last intervals are treated as open-ended with

$$\beta_0 = \sum_{m=1}^{M} w_m F_m(1)$$

and

$$\beta_{T-1} = 1 - \sum_{m=1}^{M} w_m F_m(T-1),$$

ensuring that the entry parameters sum to 1 i.e. $\sum_{b=1}^{T} \beta_{b-1} = 1$.
Figure 2 in the Supplementary Material demonstrates the modelling of the arrival process in terms of the normal mixture components and the entry parameters $\beta_{b-1}, \; b = 1, \ldots, T$.

- G: number of behavioural groups. Individuals that belong to the same group have common baseline retention probability which can be different from the corresponding probability of the other $G-1$ groups.
- $\pi_g, g = 1, \ldots, G$: The population fractions of the G behavioural groups, with $\sum_{g=1}^{G} \pi_g = 1$.
- $\phi_{gta}, g = 1, \ldots, G, t = 1, \ldots, T-1, a = t-b+1$: retention probability. The probability that a bird that belongs to behavioural group g, present at the site on occasion t and of "age" a will remain at the site until occasion $t+1$. As mentioned in Sect. 1, "age" is used to refer to the unknown time an individual has already spent at the site.
  For the particular application, retention probabilities are modelled as additive in calendar time and "age", on the logit scale, with a different intercept for each group:

$$\text{logit}(\phi_{gta}) = \gamma_{0g}^{\phi} + \gamma_1^{\phi} t + \gamma_2^{\phi} a,$$

  where $\text{logit}^{-1}(\gamma_{0g}^{\phi})$ is the baseline retention probability for behavioural group g.
- $p_t, t = 1, \ldots, T$: capture probability. The probability that a bird will be caught on occasion t given that it is present.
  For the application considered in this paper, the number of nets used and the number of hours they were left open on each capture occasion are multiplied together to form a covariate for capture probability called "effort" (e) and two dummy variables, $loc2$ and $loc3$, are created to model the effect of the three different locations where capture occasions took place during the study, in an additive logistic regression model for capture probability:

$$\text{logit}(p_t) = \gamma_0^p + \gamma_1^p e_t + \gamma_2^p \mathcal{I}(loc2_t = 1) + \gamma_3^p \mathcal{I}(loc3_t = 1),$$

where the indicator variable $\mathcal{I}(locj_t = 1)$ is 1 if capture took place on location j, $j = 2, 3$, at time t and 0 otherwise.

– $s_t, t = 1, \ldots, T$: Resighting probability. The probability that a bird will be seen on occasion t given that it is present. It is assumed to be the same for marked and unmarked birds and is modelled as constant, s, because resight occasions were conducted by the same crew which visited the same sites for the same length of time throughout the study.

The full set of parameters is

$$\boldsymbol{\theta} = \left\{ M, G, (w_m, \mu_m, \sigma_m)_{m=1,\ldots,M}, (\pi_g, \gamma_{0g}^{\phi})_{g=1,\ldots,G}, N, \gamma_1^{\phi}, \gamma_2^{\phi}, \gamma_0^p, \gamma_1^p, \gamma_2^p, \gamma_3^p, s \right\}.$$

In contrast to retention probabilities, dependence of capture and resight probabilities on "age" is not biologically meaningful and hence these parameters have not been modelled in terms of a, but if necessary such a dependence can straightforwardly be allowed for in the model. Similarly, allowing for heterogeneous groups in terms of capture/resight probabilities in the model is also possible in general, but it was not done here because of the small number of recaptures (5).

3.2 Model, prior and posterior

3.2.1 Model

Birds with CRRH h have known times of first capture, f_h, and last detection, l_h, but unknown times of arrival, b, departure, d and group membership, g. Let $\mathbf{z} = (g, b, d)$ denote the unknown life history of an individual. We can write

$$\mathcal{P}(\mathbf{z}|\boldsymbol{\theta}) = \pi_g \beta_{b-1} \left(\prod_{t=b}^{d-1} \phi_{gta} \right) (1 - \phi_{gda})^{\mathcal{I}(d<T)},$$

where the indicator variable $\mathcal{I}(d < T)$ is used to denote that the departure of individuals still present at the end of the study cannot be observed.

If $\Omega_{\mathbf{z}} = \{(g, b, d) : 1 \leq b \leq f_h \leq l_h \leq d \leq T, g \in \{1, \ldots, G\}\}$ then the probability of CRRH $\mathbf{x}_h$, $h \in \{1, \ldots, H\}$, given life history $\mathbf{z}$ and parameters in $\boldsymbol{\theta}$ is

$$\mathcal{P}(\mathbf{x}_h|\mathbf{z}, \boldsymbol{\theta}) = \begin{cases} \left[\prod_{t=b}^{d} \left\{ p_t^{\mathcal{I}(x_{ht}=1)} (1 - p_t)^{\mathcal{I}(x_{ht}=0)} \right\}^{c_t} \right] \times \\ \left[\prod_{t=f_h}^{d} \left\{ s_t^{\mathcal{I}(x_{ht}=2)} (1 - s_t)^{\mathcal{I}(x_{ht}=0)} \right\}^{r_t} \right], \mathbf{z} \in \Omega_{\mathbf{z}} \\ \\ 0, \text{otherwise} \end{cases}$$

where variable $\mathcal{I}(\psi)$ is equal to 1 if condition ψ is satisfied and 0 otherwise, $c_t = 1$ if capture took place on occasion t and variable $r_t = 1$ if instead resighting took place on occasion t, and 0 otherwise.

Similarly, if $\Omega'_{\mathbf{z}} = \{(g, b, d) : 1 \leq b \leq d \leq T, g \in \{1, \ldots, G\}\}$ the probability of the $\mathbf{0}$ history, given $\mathbf{z}, \boldsymbol{\theta}$ is

$$\mathcal{P}(\mathbf{0}|\mathbf{z}, \boldsymbol{\theta}) = \begin{cases} \left\{\prod_{t=b}^{d} (1 - p_t)^{c_t}\right\}, \mathbf{z} \in \Omega'_{\mathbf{z}} \\ \\ 0, \text{otherwise} \end{cases}.$$

Finally,

$$\mathcal{P}(\mathbf{X}, \mathbf{n}|\boldsymbol{\theta}) = \frac{N!}{\prod_h n_h!(N-D)!} \prod_h \left\{ \sum_{\mathbf{z} \in \Omega_{\mathbf{z}}} \mathcal{P}(\mathbf{z}|\boldsymbol{\theta}) \mathcal{P}(\mathbf{x}_h|\mathbf{z}, \boldsymbol{\theta}) \right\}^{n_h} \times \left\{ \sum_{\mathbf{z} \in \Omega'_{\mathbf{z}}} \mathcal{P}(\mathbf{z}|\boldsymbol{\theta}) \mathcal{P}(\mathbf{0}|\mathbf{z}, \boldsymbol{\theta}) \right\}^{N-D}.$$

M13 treated the number of unmarked birds counted on resight occasion t, y_t, as a binomially-distributed random variable with number of trials equal to N and probability of success the probability that a bird is present, unmarked and detected on that occasion. In this case, the probability of success on occasion t is

$$\zeta_t = \sum_{g=1}^{G} \sum_{b=1}^{t} \pi_g \beta_{b-1} \left(\prod_{k=b}^{t-1} \phi_{gka} \right) \left\{ \prod_{k=b}^{t} (1 - p_k)^{c_k} \right\} s_t$$

and $y_t|\boldsymbol{\theta} \sim \text{Bin}(N, \zeta_t)$. Therefore, $\mathcal{P}(\mathbf{y}|\boldsymbol{\theta}) = \prod_{t=1}^{T} \mathcal{P}(y_t|\boldsymbol{\theta})^{r_t}$, where r_t is as defined above.

3.2.2 Prior

Unless otherwise stated, simple, vague, independent priors were chosen for the model parameters. Specifically, we consider a Unif$\{1, \ldots, 20\}$ prior for M and a shifted Poisson with mean 1 for G, i.e. $G - 1 \sim \text{Po}(1)$. The principal scientific hypothesis concerning behaviour groups is that there are two, of short and long type. The natural alternative is that there is just one but it is possible that there are more than two. These prior hypotheses are represented in our Bayesian inference by a prior model over G. Most of the weight is placed on $G = 1$ and $G = 2$ but larger values of G are allowed. For M, the chosen prior reflects the lack of available information and allows for a large number of arrival groups without imposing a penalty. In fact, when $M \geq 13$, there are more parameters for modelling the entry parameters ($3 \cdot 13 = 39$) than in the M13 paper ($T = 38$). However, such large values for M, if obtained, suggest a less smooth arrival pattern that cannot be represented by a small number

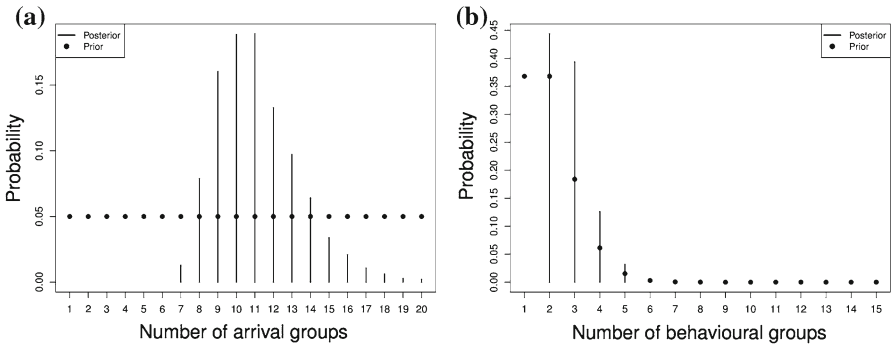

Fig. 1 Prior and posterior distribution of, **a**, M, the number of arrival groups and, **b**, G, the number of behavioural groups

of symmetric probability distributions and potentially also a lack of fit of the model. For the means of the mixture components we have chosen independent Unif$(1, T)$ to reflect our expectation that the study period encompasses the migration period while for the standard deviations we have chosen independent Unif$(0, 10)$ priors, which equally support arrival of groups spread over a period of days or in short abrupt bursts. For N we take a N$(55000, 10000^2)$ prior with the mean chosen to be close to the point estimate obtained by M13, as this reflects our expectation for the size of the super-population. $w_m, m = 1, \ldots, M$ and $\pi_g, g = 1, \ldots, G$ are given Dirichlet priors with concentration parameters all equal to 1. For the logistic regression coefficients for ϕ and p we followed Newman (2003) and King et al (2010, p. 246) who suggested mean–0 normal priors with variances equal to $\frac{\pi^2}{3(n+1)}$ where n is the number of covariates in the model. Finally, the prior for s is chosen as Beta(1, 1).

3.2.3 Posterior

Following M13,

$$\mathcal{P}(\boldsymbol{\theta}|\mathbf{X}, \mathbf{n}, \mathbf{y}) \propto \mathcal{P}(\mathbf{X}, \mathbf{n}|\boldsymbol{\theta})\mathcal{P}(\mathbf{y}|\boldsymbol{\theta})\mathcal{P}(\boldsymbol{\theta}),$$

where $\mathcal{P}(\boldsymbol{\theta})$ is the joint prior of the parameters in $\boldsymbol{\theta}$.

3.3 Results

The posterior distributions obtained for M and G are shown in Fig. 1. The first peaks at $M = 10, 11$ and sharply declines for values of $M < 9$, while its right tail is longer. The chain spent over 90 % of its time in values of $M \in \{8, \ldots, 13\}$. The latter posterior peaks at $G = 2$ and shows that the chain spent over 80 % of its time in models with $G = 2$ or $G = 3$. It gives no support to the model with $G = 1$, suggesting that there are at least two behavioural groups.

Because the values for M in $\{9, \ldots, 12\}$ are almost equally well supported, there is no clear choice for a best, or even top 2-3 models for M. Therefore, we present sum-

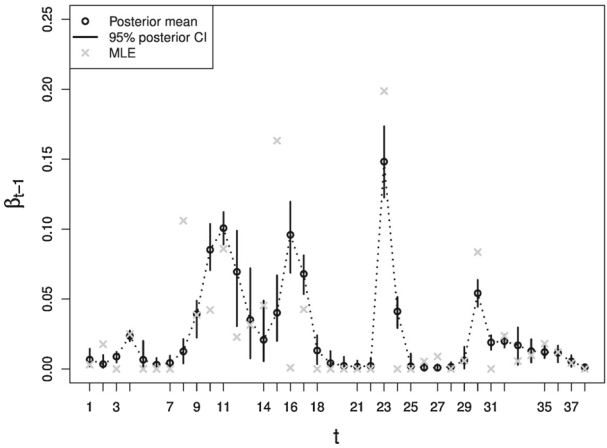

Fig. 2 Model averaged posterior means and 95% credible intervals for entry parameters, together with the maximum-likelihood estimates obtained by M13. The *tick marks* on the *x-axis* indicate days when sampling took place

maries of the model-averaged posterior distributions obtained for the entry parameters, $\beta_{t-1},\ t = 1 \ldots, T$ in Fig. 2. For comparison, we also plot the maximum-likelihood estimates obtained by M13, who estimated one entry parameter for each sample to model the arrival process. The latter are, mostly, included in the 95 % posterior credible intervals, with the exception of the point estimates obtained corresponding to the modes of the four largest peaks, which are all above the corresponding 97.5 % quantiles. Therefore, even though these two sets of estimates are not directly comparable, since a different model for the β parameters was used, they are very similar. The M13 estimates, that are only constrained to sum to 1, are more flexible but do not reflect model-averaging uncertainty. On the other hand, our model-averaged estimates provide a smoother representation of the arrival pattern of the birds at the site and are more robust to extremes. Our results suggest that the early arrival groups have greater spread in their arrival times, with arrival times overlapping between groups, while the later arrival groups are further apart and more distinct, with a longer right tail of arrivals right at the very end of the stopover period.

The posterior densities of ϕ_{gta} and π_g for $g = 1, 2$ when $G = 2$ and $g = 1, 2, 3$ when $G = 3$ when $\{t = 10, a = 1\}, \{t = 10, a = 10\}$ and $\{t = 20, a = 1\}$ are shown in Fig. 3. The areas of high density suggest two very distinct groups: a large group, with population fraction ≈80 % and low retention probability, and a small group, with population fraction ≈20 % with very high retention probability. The areas of lower density when $G = 3$ suggest that the third group, that connects the two groups, has medium retention probability. These results are consistent with the predicted differences in migration strategies of male and female sandpipers; males may spend less time than females at stopover sites in order to reach the breeding sites earlier (Bishop et al. 2004, 2006). Given the relative abundance of the retention groups in our study, this interpretation suggests a male-biased sex ratio in the stopover population. Alternatively, the heterogeneity in retention probability may reflect local movements

of birds during a period of searching and settling that often occurs immediately after arrival to a stopover area (Alerstam and Lindström 1990). The group with low retention probability may be comprised of recent arrivals that were captured during a period of searching the landscape for favourable foraging conditions, but which ultimately settled outside the study area. The smaller group with high retention probability may be comprised of birds that settled and remained in the study area during stopover. Local movements may occur in response to changing conditions in prey abundance or water depth, facilitated by wetland connectivity (Farmer and Parent 1997; Obernuefemann et al. 2013).

The effects of calendar time and "age" on retention probability are, as expected, found to be negative, with model-averaged posterior means equal to -0.645 (95% PCI: -1.032, -0.352) and -0.145 (95% PCI: -0.703, 0.317), respectively, although the effect of "age" is smaller than that of time with a credible interval that includes 0. Since the logistic regression model used to model their effects on retention probabilities is additive, plotting the joint distribution of retention probabilities and population fraction of each behavioural group for different values of time and "age" simply shifts the contours along the axis corresponding to retention probabilities, as the contour plots in Fig. 3 demonstrate.

In simulated data sets, obtained using parameter values in $\boldsymbol{\theta}$ at 100 randomly chosen simulation runs of the chain, conditional on $G = 2$, the two behavioural groups had an average observed stopover duration in days equal to 1.3 (s.d. = 0.68) and 9.95 (s.d = 7.12) while conditional on $G = 3$ the average observed stopover durations were 1.28 (s.d. = 0.62), 2.11 (s.d. = 2.18) and 10.41 (s.d. = 7.26) days.

The model-averaged mean of the posterior distribution of the super-population size, N, is equal to 62463 (95% PCI: 54412, 73151), which is greater than the estimate of M13 but with overlapping confidence bands (53595, asymptotic 95% CI: 48349, 59410). The 95% posterior credible intervals obtained for N for all combinations of the most supported values for M and G are given in Table 1 in the Supplementary Material and agree with the model-averaged interval mentioned above.

To check the fit of our specified model, we considered parameter values in $\boldsymbol{\theta}$ obtained at 500 randomly chosen simulation runs of the chain. For each of these sets of parameter values, we calculated the value of the log-likelihood for the real data set and for a data set simulated from this set. The distributions of these two sets of log-likelihood values, plotted in Fig. 18 in the Supplementary Material, peak at the same point, which suggests a good fit of the model. For each of these simulated data sets we also obtained the number of birds first caught on each capture occasion and the number of birds resighted as unmarked on each resighting occasion. Figure 18 in the Supplementary Material, respectively, show that the observed values almost always overlap with the boxplots of these simulated values, once more providing support to the claim that the model fits well. Finally, for each of these simulated data sets we obtained the total number of individuals caught at least once and we compared it to the value observed for the real data set ($D = 507$). Hence, we used the values for D as our test statistic in our calculation of our Bayesian p-value, that is the proportion of time the simulated values for D are less than the value of D obtained for the real data set, which being equal to 0.51 does not raise any concerns about the fit of the model.

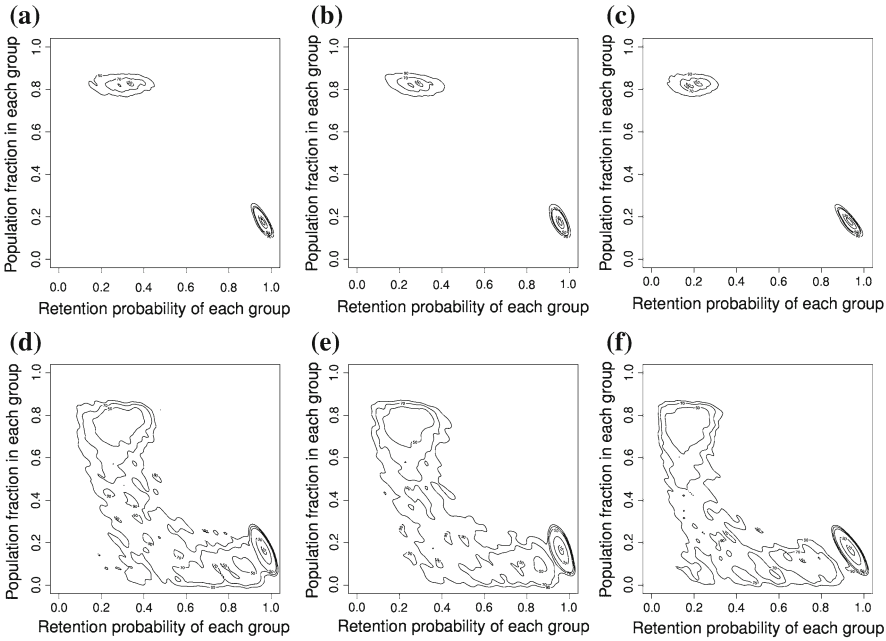

Fig. 3 Contour plots of the joint posterior densities for retention probability and population fraction of each behavioural group when $G = 2$ (*top row*) and $G = 3$ (*bottom row*), when time $t = 10$ and "age" $a = 1$ (*first column*), time $t = 10$ and "age" $a = 10$ (*middle column*), time $t = 20$ and "age" $a = 1$ (*last column*). These are constructed by ordering the groups during post-processing in terms of their baseline retention probability, obtaining the pairs of proportions and retention probabilities sampled at each iteration, conditional on G, and calculating the kernel density estimate using R package ks (Duong 2016). The figure which also includes the scatterplot of these pairs of values for each group is shown as Supplementary Material

4 Discussion

To analyse the stopover data set considered in Sect. 3 we extended existing stopover models to allow for individuals to arrive in groups and to exhibit heterogeneity in their stopover duration at the site. We showed that both these processes can be modelled at the same time and the uncertainty in the number of groups in either process can be accounted for using a RJMCMC algorithm. Our results suggest that semipalmated sandpipers at stopover sites do not exhibit the same behaviour in terms of their stopover.

By using finite mixtures of continuous distributions, such as normal, to model the arrival of individuals at the study site, instead of models with fully time-dependent entry probabilities as in M13, the number of parameters does not necessarily increase with the number of sampling occasions. Additionally, one is supplied with an uncomplicated and biologically meaningful way to interpret the results in terms of the number of arrival groups and their behaviour, making analyses on different data sets, for example from different years, directly comparable. Moreover, the parameters of the mixture components, such as the mean arrival times, can also be modelled as functions of, for example, weather covariates. Further simplifications of the models are also possible.

Specifically, it might be assumed that the means of the arrival times of the different groups are equally spaced, with the space to be estimated by the model. We have considered the case of normal mixtures for the work in this paper but other distributions, not necessarily symmetric, could also be chosen, such as gamma, if appropriate.

Bayesian inference enables the incorporation of prior beliefs which might not treat all models as equally likely a priori. The belief that there are two behavioural group of birds at the stopover site, namely the short– and the long–stayers was easily incorporated in the model, instead of naively assuming that a model with 15 behavioural groups is as likely as the more realistic one with just two groups. On the other hand, posterior model probabilities are known to be sensitive to prior model probabilities (Corani and Mignatti 2015) and hence, the latter should be chosen very carefully.

The application presented demonstrates the general applicability of the RJMCMC algorithm, even when the population is heterogeneous in more than one process, for instance both in survival and arrival and the models are highly complex. Unaccounted-for hererogeneity can lead to biased parameter estimates and spurious results. Specifically, it has been frequently reported that unmodelled heterogeneity in capture probabilities leads to biased estimates of the population size (Pollock et al. 1990), but it can also affect estimation of survival probabilities (Oliver et al. 2011; Fletcher et al. 2012; Matechou et al. 2013b). If potential heterogeneity in survival probabilities remains unmodelled, then individuals with an overall higher survival probability will prevail at older ages, which can result in the average survival probability appearing to increase by age (Vaupel and Yashin 1985; Peron et al. 2010), masking the effect of senescence. Accounting for heterogeneity is also important in non-ecological applications of CR models with an emphasis on estimating population size (see McCrea and Morgan 2014, p. 46).

Even when the list of possible models to be considered is large, as in the example of Sect. 3, the use of the RJ algorithm makes model selection possible since inappropriate models do not have to be actually fitted, i.e. visited by the algorithm. When appropriate, the use of RJMCMC enables model-averaging and does not require the quite often unclear or subjective choice of one single "best" model. The posterior density obtained for the number of arrival groups gives very similar support to $M = 9, \ldots, 12$ groups, and the conclusions have been drawn by averaging over these, as well as the less supported models.

We have not considered heterogeneity in capture probabilities for the population of semipalmated sandpipers because of the low number of recaptures and we suggested in Sect. 3 that the models and the RJMCMC algorithm can be extended to that effect. However, it should be noted here that, as Link (2003) explains and Arnold et al. (2010) discuss, parameter N is not identifiable among different model classes, for example between finite and infinite mixture models, and it may not be possible to distinguish between models with different assumptions about capture probability that provide very different estimates for N. Additionally, the typical sparseness of ecological data sets may limit the complexity of the model and introduce identifiability issues if heterogeneity were to be considered in both retention and capture/resighting probabilities.

We demonstrated the models in this paper by considering a data set of migrating semipalmated sandpipers collected at a stopover site. However, the models are more

generally applicable as other species and animals arrive or emerge in groups and exhibit heterogeneity in their survival or detection. For example they could apply to data sets of amphibians collected at breeding ponds, with different groups expected to arrive at different times (Harrison et al. 2009, observed male newts arriving before females), and in addition to have different detection and/or retention probabilities.

Acknowledgments We are grateful to François Caron for his suggestions and comments. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. The findings and conclusions in this article are those of the authors and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Caski F (eds) Proceeding of the Second International Symposium on Information Theory Akademiai Kiado, Budapest pp 267–281

Alerstam T, Lindström A (1990) Optimal migration: the relative importance of time, energy, and safety. In: Gwinner E (ed) Bird migration: physiology and ecophysiology. Springer-Verlag, Berlin, pp 331–351

Arnold R, Hayakawa Y, Yip P (2010) Capture-recapture estimation using finite mixtures of arbitrary dimension. Biometrics 66(2):644–655

Besbeas P, Freeman SN, Morgan BJT, Catchpole EA (2002) Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. Biometrics 58:540–547

Biernacki C, Celeux G, Govaert G (2000) Assessing a mixture model for clustering with the integrated completed likelihood. IEEE Trans Pattern Anal Mach Intell 22(7):719–725

Bishop MA, Warnock N, Takekawa JY (2004) Differential spring migration by male and female western sandpipers at interior and coastal stopover sites. Ardea 92(2):185–196

Bishop MA, Warnock N, Takekawa JY (2006) Spring migration patterns in Western sandpipers Calidris mauri. In: Boere GC, Galbraith CA, Stroud DA (eds) Waterbirds around the world. The Stationery Office, Edinburgh

Brooks SP, Catchpole EA, Morgan BJT (2000) Bayesian animal survival estimation. Stat Sci 15(4):357–376

Brown S, Hickey C, Harrington B, Gill R (2001) The US shorebird conservation plan, 2nd edn. Manomet Center for Conservation Sciences, Manomet

Burnham KP, Anderson DR (2002) Model selection and multimodel inference: a practical information-theorerical approach. Springer-Verlag, New York

Corani G, Mignatti A (2015) Robust Bayesian model averaging for the analysis of presence–absence data. Environ Ecol Stat 22(3):513–534

Cristol DA, Baker MB, Carbone C (1999) Differential migration revisited. In: Nolan VJ, Ketterson ED, Thompson CF (eds) Current ornithology, vol 15. Springer, US, pp 33–88

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

Dinsmore SJ, Collazo JA (2003) The influence of body condition on local apparent survival of spring migrant sanderlings in coastal North Carolina. The Condor 105(3):465–473

Duong T (2016) ks: Kernel Smoothing. https://CRAN.R-project.org/package=ks, r package version 1.10.3

Farmer AH, Parent AH (1997) Effects of the landscape on shorebird movements at spring migration stopovers. The Condor 99:697–707

Fletcher D, Lebreton JD, Marescot L, Schaub M, Gimenez O, Dawson S, Slooten E (2012) Bias in estimation of adult survival and asymptotic population growth rate caused by undetected capture heterogeneity. Methods Ecol Evol 3:206–216

Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. Biometrika 82:711–732

Harrison JD, Gittins SP, Slater FM (2009) The breeding migration of smooth and Palmate newts (Triturus vulgaris and T. helveticus) at a pond in mid Wales. J Zool 199:249–258. doi:10.1111/j.1469-7998.1983.tb02093.x

Jolly GM (1965) Explicit estimates from capture-recapture data with both death and immigration-stochastic model. Biometrika 52:225–247

King R, Brooks SP (2008) On the Bayesian estimation of a closed population size in the presence of heterogeneity and model uncertainty. Biometrics 64:816–824

King R, Morgan BJT, Gimenez O, Brooks SP (2010) Bayesian analysis for population ecology. CRC Press, Boca Raton

Link WA (2003) Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. Biometrics 59(4):1123–1130

Lyons JE, Haig SM (1995) Fat content and stopover ecology of spring migrant semipalmated sandpipers in South Carolina. The Condor 97:427–437

Lyons JE, Runge MC, Laskowski HP, Kendall WL (2008) Monitoring in the context of structured decision-making and adaptive management. J Wildl Manag 72:1683–1692

Matechou E, Morgan BJT, Pledger S, Collazo JA, Lyons JE (2013a) Integrated analysis of capture-recapture-resighting data and counts of unmarked birds at stop-over sites. J Agric Biol Environ Stat 18:120–135

Matechou E, Pledger S, Efford M, Morgan BJT, Thomson DL (2013b) Estimating age-specific survival when age is unknown: open population capture-recapture models with age structure and heterogeneity. Methods Ecol Evol 4:654–664

Matechou E, Dennis EB, Freeman SN, Brereton T (2014) Monitoring abundance and phenology in (multi-voltine) butterfly species: a novel mixture model. J Appl Ecol 51:766–775

McCrea RS, Morgan BJT (2014) Analysis of capture–recapture data. CNC Press, Boca Raton

McLachlan G, Peel D (2000) Finite mixture models. Wiley, New Jersey

Newman KB (2003) Modelling paired release-recovery data in the presence of survival and capture heterogeneity with application to marked juvenile salmon. Statistical modelling 3:157–177

Nichols JD, Williams BK (2006) Monitoring for conservation. Trends Ecol Evol 21:668–673

Obernuefemann KP, Collazo JA, Lyons JE (2013) Local movements and wetland connectivity at a migratory stopover of semipalmated sandpipers in southeastern United States. Waterbirds 36:62–74

Oliver LJ, Morgan BJT, Durant SM, Pettorelli N (2011) Individual heterogeneity in recapture probability and survival estimates in cheetah. Ecol Model 222:776–784

Peron G, Crochet PA, Choquet R, Pradel R, Lebreton JD, Gimenez O (2010) Capture-recapture models with heterogeneity to study survival senescence in the wild. Oikos 119:524–532

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

Pledger S, Efford M, Pollock KH, Collazo JA, Lyons JE (2009) Stopover duration analysis with departure probability dependent on unknown time since arrival. Environmental and ecological statistics (Edited by DLThomson, EGCooch and MJ Conroy) 3:349–363

Pledger S, Pollock KH, Norris JL (2010) Open capture-recapture models with heterogeneity: II. Jolly–Seber model. Biometrics 66:883–890

Plummer M, Best N, Cowles K, Vines K (2006) Coda: convergence diagnosis and output analysis for MCMC. R News 6(1):7–11, http://CRAN.R-project.org/doc/Rnews/

Pollock KH, Nichols JD, Brownie C, Hines JE (1990) Statistical inference for capture-recapture experiments. Wildl Monogr 107:3–97

R Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/

Rubolini D, Spina F, Saino N (2004) Protandry and sexual dimorphism in trans-saharan migratory birds. Behav Ecol 15(4):592–601

Schwarz CJ, Arnason AN (1996) A general methodology for the analysis of capture-recapture experiments in open populations. Biometrics 52:860–873

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Seber GAF (1965) A note on the multiple-recapture census. Biometrika 52:249–259

Tavaré S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from dna sequence data. Genetics 145(2):505–518

Vaupel JW, Yashin AI (1985) Heterogeneity's ruses: some surprising effects of selection on population dynamics. The Am Statistician 39:176–185

Eleni Matechou obtained her Ph.D. in Statistics from the University of Kent in 2010. She spent 1 academic year at Victoria University in Wellington as a Research and Teaching Fellow and 3 years at Oxford University as a Departmental Lecturer and Stipendiary Lecturer for St. Peter's College. She is now employed as a Lecturer in Statistics at the University of Kent. She is a member of the National Centre for Statistical Ecology (http://ncse.org.uk/) and a committee member of the Environmental Statistics Section of the Royal Statistical Society (http://tinyurl.com/lkzlkhy). She is generally interested in developing population ecology models for open populations and she have published papers on models for capture-recapture data as well as on models for count data of butterflies resulting from citizen science surveys.

Geoff K. Nicholls is a Fellow of St Peter's college Oxford and an Associate Professor in the Statistics Department at Oxford. Geoff completed his Ph.D. in High Energy Particle Physics in the Department of Applied Mathematics and Theoretical Physics in Cambridge University in 1990. He was a research assistant 1990–1992 in the Vision Speech and Signal Processing group in the Department of Electronic and Electrical Engineering at the University of Surrey in Guildford, and again 1992–1995 in the Statistics Department at Oxford University. He was for ten years 1995–2005 a lecturer in the Mathematics Department of the University of Auckland in New Zealand before moving to his present post in 2005. Geoff was Head of Department in Statistics at Oxford from 2012 to 2015. He is interested in statistical methodology, applied Bayesian statistics and computational problems associated with building and fitting complex stochastic models.

Byron J. T. Morgan is Emeritus Professor and Honorary Professorial Research Fellow in the School of Mathematics, Statistics & Actuarial Science at the University of Kent, England. He is the author/coauthor of books and papers in a number of different areas, including bioassay, capturerecapture, parameter redundancy, population ecology and simulation. He is Co-Director of the UK National Centre for Statistical Ecology.

Jaime A. Collazo obtained his Ph.D. in Animal Ecology from Iowa State University in 1985. He joined the faculty at North Carolina State in 1989, and currently he is a Professor in Applied Ecology and Assistant Leader of the North Carolina Cooperative Fish and Wildlife Research Unit. Research interests are in the realm of conservation, with emphasis on avian population dynamic, species-habitat relationships, and endangered species.

James E. Lyons obtained his Ph.D in Biology from Virginia Tech University, Blacksburg, VA, USA and M.S. in Wildlife Biology from Clemson University, Clemson, SC, USA. Following postdoctoral research at North Carolina State University and Patuxent Wildlife Research Center, he worked as a wildlife biologist in the Migratory Bird Program of the U.S. Fish and Wildlife Service for nine years. He is currently a research ecologist with U.S. Geological Survey and focus on the application of hierarchical models and decision science for conservation and management of wildlife populations.