Efficient sequential Monte Carlo algorithms for integrated population models

Axel Finke[1] axel.finke@nus.edu.sg
Ruth King[3,4] ruth.king@ed.ac.uk
Alexandros Beskos[2,4] a.beskos@ucl.ac.uk
Petros Dellaportas[2,4,5] p.dellaportas@ucl.ac.uk

[1]Department of Statistics and Applied Probability, National University of Singapore, Singapore
[2]Department of Statistical Science, University College London, U.K.
[3]School of Mathematics, University of Edinburgh, U.K.
[4]The Alan Turing Institute, U.K.
[5]Department of Statistics, Athens University of Economics and Business, Greece

Abstract

In statistical ecology, state-space models are commonly used to represent the biological mechanisms by which population counts – often subdivided according to characteristics such as age group, gender or breeding status – evolve over time. As the counts are only noisily or partially observed, they are typically not sufficiently informative about demographic parameters of interest and must be combined with additional ecological observations within an integrated data analysis. Fitting integrated models can be challenging, especially if the constituent state-space model is non-linear/non-Gaussian. We first propose an efficient particle Markov chain Monte Carlo algorithm to estimate demographic parameters without a need for linear or Gaussian approximations. We then incorporate this algorithm into a sequential Monte Carlo sampler to perform model comparison. We also exploit the integrated model structure to enhance the efficiency of both algorithms. The methods are demonstrated on two real data sets: little owls and grey herons. For the owls, we find that the data do not support an ecological hypothesis found in the literature. For the herons, our methodology highlights the limitations of existing models which we address through a novel regime-switching model.

Key words: Bayesian inference; Capture-recapture; Integrated population models; Model comparison; Sequential Monte Carlo; State-space models.

1 Introduction

State-space models are an increasingly common and useful representation of many ecological systems (Buckland et al., 2007; King, 2014; Newman et al., 2014). They are used to describe, e.g., population count data (Newman, 1998; Besbeas et al., 2002; King et al., 2008); telemetry data (Morales et al., 2004; McClintock et al., 2012; Breed et al., 2012); longitudinal growth data (Peters et al., 2010); fisheries biomass dynamics (Millar and Meyer, 2000) or capture-recapture data (Dupuis, 1995; Royle, 2008; King, 2012).

Inference in state-space models. Fitting state-space models is computationally challenging as the likelihood – expressible as an integral over unobserved states – is typically intractable unless the states take values in a small, finite set; or the model is linear-Gaussian whence the likelihood is evaluated via the Kalman filter (Kalman, 1960; Newman, 1998). Two approaches are typically applied to circumvent this problem. The first is to approximate the state-space model with one that is linear and Gaussian (Besbeas et al., 2002) but this introduces a difficult-to-quantify bias. The second is to impute the unobserved states alongside the model parameters within an Markov chain Monte Carlo (MCMC) algorithm. Such data-augmentation schemes – in particular as implemented in *BUGS* (Gilks et al., 1994) or *JAGS* (Plummer, 2003) (see e.g. Brooks et al., 2004) – can be slow and poorly mixing if the states and parameters are considerably correlated because only (small) subsets of them are updated individually (King, 2011). To avoid these problems, Andrieu et al. (2010) proposed particle Markov chain Monte Carlo (PMCMC) algorithms (see Knape and de Valpine, 2012; Parslow et al., 2013, for recent applications in ecology). Such algorithms, implemented e.g. in the software *Nimble* (de Valpine et al., 2017), replace the intractable likelihood in the MCMC acceptance ratio with an estimate obtained through a sequential Monte Carlo (SMC) algorithm (or "particle filter") without introducing bias. Another popular approach, implemented in the software *Stan* (Carpenter et al., 2017), simultaneously explores multiple directions of the space via Hamiltonian dynamics. However, as these rely on the gradient of the log-posterior, they are primarily designed for continuous latent variables (and the models treated in

this work have discrete latent variables); extensions to discrete spaces have recently been attempted (Nishimura et al., 2017) but these are still far from established.

Integrated population models. In this work, count data are available on some species of interest, i.e. estimates of population sizes at discrete times (Besbeas et al., 2002, 2009; King et al., 2008). These count data are modelled as a state-space model to account for measurement errors. In addition to the count data, other types of data are available on the species, e.g. capture-recapture, ring-recovery or nest-record data. To utilise all available information for estimating demographic parameters of interest, it is necessary to combine these different data sets within a single *integrated population model*. Unfortunately, fitting such models is challenging, since they inherit all the above-mentioned difficulties with fitting the constituent state-space model.

Contributions. In this work, we devise efficient methodology for performing fully Bayesian parameter estimation and model comparison in integrated population models without the need for linear or Gaussian approximations to the state-space model.

- In Section 3, we first review standard PMCMC methods for Bayesian parameter estimation in models with intractable likelihoods. Then, in Subsection 3.2, we exploit the integrated model structure to reduce the computational cost of the PMCMC algorithm through a delayed-acceptance (Christen and Fox, 2005) technique.

- In Section 4, we first incorporate our PMCMC methodology into SMC samplers (Chopin, 2002; Del Moral et al., 2006; Duan and Fulop, 2015; Zhou et al., 2016) so that we can estimate posterior model probabilities (Bayes factors) across a set of different integrated population models. This permits Bayesian model comparison without the need for reversible-jump MCMC (Green, 1995) which often mixes poorly and can be difficult to implement and tune. Then, in Subsection 4.3, we again exploit the integrated-model structure to reduce the computational burden of the SMC sampler by separately tempering the different likelihood terms.

- In Sections 5 and 6, we apply the proposed methodology to two real data sets relating to little owls and grey herons and obtain estimates of the evidence for a number

of models proposed in the literature. For the owls, we find that the model from Abadi et al. (2010b) may be unnecessarily over-parametrised; in particular, we find no evidence for the hypothesis in Abadi et al. (2010b) that the owls' immigration rate depends on the abundance of voles – their main prey. We also demonstrate the utility of the delayed-acceptance approach. For the herons, we show that an elaborate threshold model from Besbeas and Morgan (2012) fits poorly; to remedy this, we propose a novel regime-switching state-space model which significantly outperforms all existing models in terms of model fit and model evidence.

2 Integrated model

2.1 Data

We combine multiple data sets, one of which being count data, obtained from a single population, within a single integrated model. Let $\mathbf{y} = \{y_1, \ldots, y_T\}$ denote *count data* collected at times $t = 1, \ldots, T$. Here, y_t is the observation (subject to measurement error) of the true population size at time t. The observed counts may be multivariate, e.g. counts for males and females or juveniles and adults, though in all the examples we consider later the count data are univariate. Let $\mathbf{w}$ denote all *additional data* available such as capture-recapture data, ring-recovery or nest-record data. The aim of this work is then to perform inference based on *all data* $\mathbf{z} = \{\mathbf{y}, \mathbf{w}\}$. To illustrate the methodology, we consider two data sets relating to little owls and grey herons.

Little owls. In Section 5, we consider little owl data described by Schaub et al. (2006) and subsequently analysed in Abadi et al. (2010b). The count data represent the number of breeding females at nest boxes near Göppingen, South Germany, observed annually from 1978 to 2003 (i.e. $T = 26$). The nest boxes were checked multiple times annually and data were recorded relating to overall population size (numbers of occupied nest boxes and breeding females), capture-recapture histories of individuals observed at nest boxes and reproductive success of the nests. In addition, time-varying covariate information about the abundance of voles – the primary prey for little owls – is available.

For further details see Schaub et al. (2006).

Grey herons. In Section 6, we consider grey-heron data previously presented and analysed by Besbeas et al. (2002, 2009) and Besbeas and Morgan (2012). The count data correspond to the estimated number of female herons (or breeding pairs) in the UK, from 1928 to 1998, i.e. for $T = 71$ time periods. Within our application we also have ring-recovery data for individuals released between 1955 and 1997.

2.2 Model structure

Given unknown model parameters $\theta \in \Theta$, the likelihood of the count data $\mathbf{y}$ and additional data $\mathbf{w}$ is $p(\mathbf{z}|\theta) = p(\mathbf{y}|\theta, \mathbf{w})p(\mathbf{w}|\theta)$. To simplify the notation, and in agreement with ecological practice (see e.g. Besbeas et al., 2002), we assume that the count data are independent of the additional data given the parameters, i.e. $p(\mathbf{y}|\theta, \mathbf{w}) = p(\mathbf{y}|\theta)$ (see Abadi et al., 2010a, for a justification). However, we stress that this conditional independence is purely a modelling choice; our methodology only requires that the additional data are modelled in such a way that $p(\mathbf{w}|\theta)$ can be evaluated pointwise.

The count data are described by a state-space model as follows. Let $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_T\} \in \mathsf{X}^T$ (for some space X) denote the true (unobserved) population counts with initial density $\mu_\theta(\mathbf{x}_1)$ and transitions $f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1})$. Furthermore, let $g_\theta(y_t|\mathbf{x}_t)$ be the density of the tth observed count given $\mathbf{x}_t$. Then, conditionally on θ, the joint distribution of $\mathbf{y}$ and $\mathbf{x}$ is:

$$p(\mathbf{y}, \mathbf{x}|\theta) = \mu_\theta(\mathbf{x}_1) g_\theta(y_1|\mathbf{x}_1) \prod_{t=2}^{T} f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1}) g_\theta(y_t|\mathbf{x}_t).$$

The (marginal) count-data likelihood is thus given by the integral (or sum, if X is discrete)

$$p(\mathbf{y}|\theta) = \int_{\mathsf{X}^T} p(\mathbf{y}, \mathbf{x}|\theta)\, \mathrm{d}\mathbf{x}. \tag{1}$$

Throughout this work, we assume that this integral (sum) is intractable as is usually the case unless X is finite and sufficiently small or the state-space model is linear and Gaussian in which case (1) can be evaluated using the Kalman filter.

Let $p(\theta)$ denote the prior distribution of the parameters then the (marginal) *posterior*

distribution of the parameters θ (given the full data $\mathbf{z}$) is given by

$$\pi(\theta) \coloneqq p(\theta|\mathbf{z}) = \frac{p(\mathbf{z}|\theta)p(\theta)}{p(\mathbf{z})}; \quad p(\mathbf{z}) \coloneqq \int_{\Theta} p(\mathbf{z}|\theta)p(\theta)\mathrm{d}\theta, \tag{2}$$

where $p(\mathbf{z})$ in the denominator is the *evidence* for the model. This quantity plays a key role in Bayesian model comparison as outlined in Section 4. The posterior distribution is typically intractable as the integrals in (1), (2) are not of closed form. Instead, we approximate it via Monte Carlo methods as described in the next section.

3 Parameter estimation

3.1 Particle MCMC

In this section, we describe MCMC methods for approximating the posterior distribution of the model parameters. We also propose modifications which exploit the structure of integrated models to improve efficiency of the algorithm. For now, we assume that the model is known – model uncertainty is dealt with in Section 4.

As the count-data likelihood $p(\mathbf{y}|\theta)$ (and thus $p(\mathbf{z}|\theta)$) is intractable, we cannot implement the *idealised* Metropolis–Hastings algorithm which, at each iteration, proposes new parameters $\vartheta \sim q(\vartheta|\theta)$ and accepts them with probability (w.p.) $1 \wedge \frac{q(\theta|\vartheta)}{q(\vartheta|\theta)} \frac{p(\vartheta)}{p(\theta)} \frac{p(\mathbf{z}|\vartheta)}{p(\mathbf{z}|\theta)}$. A common solution is to use data-augmentation, i.e. to impute the latent variables $\mathbf{x}$ (alongside the parameters). However, the number of states is typically large so that single-site updates (i.e. updates for a single state $\mathbf{x}_t$ conditional on $\{\mathbf{x}_s \colon s \neq t\}$ within a Gibbs sampler framework) are required. This approach, commonly used in 'black-box' samplers such as *BUGS* or *JAGS*, can lead to poor mixing if highly correlated variables or parameters are updated separately. To avoid such problems, we employ particle Markov chain Monte Carlo (PMCMC) algorithms (Andrieu et al., 2010). These replace $p(\mathbf{y}|\theta)$ in the acceptance ratio of the idealised Metropolis–Hastings algorithm with an unbiased estimate $\hat{p}(\mathbf{y}|\theta)$ obtained through a sequential Monte Carlo (SMC) method. Crucially, the resulting algorithm still targets the correct posterior distribution.

Before stating the PMCMC algorithm, we review SMC algorithms, usually termed

particle filters (PFs) when applied to state-space models (Doucet and Johansen, 2011). A simple PF is outlined in Algorithm 1, where we use the convention that actions prescribed for the nth particle are to be performed independently for all $1 \leq n \leq N$, for some user-specified number, $N \geq 1$, of particles.

---

1 Algorithm (particle filter). Sample $\mathbf{x}_t^n \sim \mu_\theta(\mathbf{x}_t)$; for $t = 2, \ldots, T$,

(1) sample $a_{t-1}^n = l \in \{1, \ldots, N\}$ w.p. $W_{t-1}^l \propto w_{t-1}^l \coloneqq g_\theta(y_{t-1}|\mathbf{x}_{t-1}^l)$,

(2) sample $\mathbf{x}_t^n \sim f_\theta(\mathbf{x}_t|\mathbf{x}_{t-1}^{a_{t-1}^n})$.

At the end of Algorithm 1, an unbiased (Del Moral, 1996) estimate of $p(\mathbf{y}|\theta)$ is

$$\hat{p}(\mathbf{y}|\theta) \coloneqq \prod_{t=1}^{T} \frac{1}{N} \sum_{n=1}^{N} w_t^n.$$

Numerous extensions exist for making Algorithm 1 more efficient. The particular version of PF we use in our applications is outlined in Web Appendix C.

We now describe the PMCMC algorithm. A single PMCMC update is outlined in Algorithm 2, where $\alpha \in [0, 1]$ is a parameter which will be used by the evidence-approximation algorithms in Section 4. For the moment simply consider $\alpha = 1$.

---

2 Algorithm (particle MCMC). At each iteration, given $(\theta, \hat{p}(\mathbf{y}|\theta))$,

(1) propose $\vartheta \sim q(\vartheta|\theta)$ and generate $\hat{p}(\mathbf{y}|\vartheta)$ using Alg. 1 (wherein $\theta = \vartheta$),

(2) return $(\vartheta, \hat{p}(\mathbf{y}|\vartheta))$ w.p. $1 \wedge \dfrac{q(\theta|\vartheta)}{q(\vartheta|\theta)} \dfrac{p(\vartheta)}{p(\theta)} \left[\dfrac{\hat{p}(\mathbf{y}|\vartheta)p(\mathbf{w}|\vartheta)}{\hat{p}(\mathbf{y}|\theta)p(\mathbf{w}|\theta)}\right]^\alpha$; otherwise, return $(\theta, \hat{p}(\mathbf{y}|\theta))$.

3.2 Improving PMCMC efficiency for integrated models

The computational cost of the PMCMC update in Algorithm 2 is dominated by the PF used to generate the estimate of $p(\mathbf{y}|\vartheta)$ for each proposed parameter value ϑ. To improve the efficiency of the algorithm, we utilise a delayed-acceptance (DA) approach (Christen and Fox, 2005; Sherlock et al., 2015) based on the integrated-model structure. The idea is to avoid invoking the PF for proposed parameter values ϑ which are not compatible with the additional data $\mathbf{w}$ and are therefore likely to be rejected in Algorithm 2. This

can improve efficiency if $\mathbf{w}$ is highly informative about a considerable proportion of the parameters. Algorithm 3 summarises the approach. Again, take $\alpha = 1$, for the moment.

---

3 Algorithm (delayed acceptance PMCMC). At each iteration, given $(\theta, \hat{p}(\mathbf{y}|\theta))$:

(1) Propose $\vartheta \sim q(\vartheta|\theta)$,

(2) Go to Step 3 w.p. $1 \wedge \dfrac{q(\theta|\vartheta)}{q(\vartheta|\theta)} \dfrac{p(\vartheta)}{p(\theta)} \left[\dfrac{p(\mathbf{w}|\vartheta)}{p(\mathbf{w}|\theta)}\right]^{\alpha}$; otherwise, return $(\theta, \hat{p}(\mathbf{y}|\theta))$.

(3) Generate $\hat{p}(\mathbf{y}|\vartheta)$ using Alg. 1 (with $\theta = \vartheta$).

(4) Return $(\vartheta, \hat{p}(\mathbf{y}|\vartheta))$ w.p. $1 \wedge \left[\dfrac{\hat{p}(\mathbf{y}|\vartheta)}{\hat{p}(\mathbf{y}|\theta)}\right]^{\alpha}$; otherwise, return $(\theta, \hat{p}(\mathbf{y}|\theta))$.

The validity of Algorithm 3 may be established using the arguments of Christen and Fox (2005), Andrieu et al. (2010). We note DA was previously combined with PMCMC updates in Golightly et al. (2015) (though in a different way).

4 Model comparison

4.1 Posterior model probabilities

Let $\{\mathcal{M}_i \colon 1 \leq i \leq I\}$ denote a finite collection of plausible biological models. To indicate the ith model, we add the model indicator $\mathcal{M}_i$ to densities in Section 2 so that the prior of $\theta \in \Theta_i$ is written as $p(\theta|\mathcal{M}_i)$, the likelihood as $p(\mathbf{z}|\theta, \mathcal{M}_i) = p(\mathbf{y}|\theta, \mathcal{M}_i)p(\mathbf{w}|\theta, \mathcal{M}_i)$ and the evidence as $p(\mathbf{z}|\mathcal{M}_i) = \int_{\Theta_i} p(\mathbf{z}|\theta, \mathcal{M}_i)p(\theta|\mathcal{M}_i)\,\mathrm{d}\theta$. Let $p(\mathcal{M}_i)$ denote the prior probability of the ith model. Bayesian model comparison is based on the *posterior model probabilities* (Bernardo and Smith, 2009, Chapter 6)

$$p(\mathcal{M}_i|\mathbf{z}) \coloneqq \frac{p(\mathcal{M}_i)p(\mathbf{z}|\mathcal{M}_i)}{\sum_{j=1}^{I} p(\mathcal{M}_j)p(\mathbf{z}|\mathcal{M}_j)}. \tag{3}$$

Unfortunately, the model evidence $p(\mathbf{z}|\mathcal{M}_i)$ in (3) – hence also ith posterior model probability – is intractable. To perform model comparison, we replace the model evidence $p(\mathbf{z}|\mathcal{M}_i)$ with an estimate $\hat{p}(\mathbf{z}|\mathcal{M}_i)$ obtained via an SMC sampler. As a by-product, the SMC sampler also yields an approximation of the posterior of θ under the ith model.

4.2 SMC sampler for evidence approximation

For the moment, assume that $p(\mathbf{y}|\theta, \mathcal{M}_i)$ can be evaluated. A simple importance-sampling approximation of $p(\mathbf{z}|\mathcal{M}_i)$ is then given by $\frac{1}{M}\sum_{m=1}^{M} p(\mathbf{z}|\theta^m, \mathcal{M}_i)$, where $\theta^1, \ldots, \theta^M$ are sampled independently from $p(\theta|\mathcal{M}_i)$. However, a prohibitively large sample size M is required if there is a strong mismatch between the prior and the posterior (which is common, especially if θ is high-dimensional or if the data are highly informative). To circumvent this problem, we employ an SMC sampler (Chopin, 2002; Del Moral et al., 2006) which uses successive importance-sampling steps to approximate a *sequence* of distributions to smoothly bridge the gap between the prior and the posterior,

$$p(\theta|\mathcal{M}_i) = \pi_0(\theta), \pi_1(\theta), \ldots, \pi_S(\theta) = p(\theta|\mathbf{z}, \mathcal{M}_i). \tag{4}$$

The idea behind SMC samplers is that each individual importance-sampling step (i.e. proposing samples from $\pi_{s-1}(\theta)$ to approximate $\pi_s(\theta)$) may be feasible even if the gap between prior $\pi_0(\theta)$ and posterior $\pi_S(\theta)$ is wide. We use a likelihood-tempering approach,

$$\pi_s(\theta) \propto p(\theta|\mathcal{M}_i)p(\mathbf{z}|\theta, \mathcal{M}_i)^{\alpha_s}, \tag{5}$$

where the *temperatures* $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_S = 1$ (and the number of bridging distributions, S) can then be tuned to ensure that the interpolation between the prior and posterior in (4) is sufficiently smooth. In the models considered in this work, $p(\mathbf{y}|\theta)$ is intractable and therefore again approximated using a PF (for any $0 < \alpha_s < 1$, the distributions targeted by the algorithm are then actually slightly different from (5) but we stress that this does not affect the validity of the algorithm). This idea was first employed by Duan and Fulop (2015).

Algorithm 4 outlines the SMC sampler; we use the convention that any action specified for the mth particle is to be performed independently for *all* $m \in \{1, \ldots, M\}$.

---

4 Algorithm (SMC sampler).

(1) Sample $\theta_0^m \sim p(\theta|\mathcal{M}_i)$ and generate $\hat{p}_0^m(\mathbf{y}|\theta_0^m, \mathcal{M}_i)$ using Alg. 1 (with $\theta = \theta_0^m$),

(2) At step $s = 1, \ldots, S$,

(a) set $v_s^m \coloneqq (u_{s-1}^m)^{\alpha_s - \alpha_{s-1}}$, where $u_{s-1}^m \coloneqq \hat{p}_{s-1}^m(\mathbf{y}|\theta_{s-1}^m, \mathcal{M}_i)p(\mathbf{w}|\theta_{s-1}^m, \mathcal{M}_i)$,

(b) sample $b_{s-1}^m = l \in \{1, \ldots, M\}$ w.p. $V_s^l \propto v_s^l$,

(c) sample $(\theta_s^m, \hat{p}_s(\mathbf{y}|\theta_s^m, \mathcal{M}_i))$ using Alg. 2 (with $\alpha = \alpha_s$; $\theta = \theta_{s-1}^{b_{s-1}^m}$; $\hat{p}(\mathbf{y}|\theta) = \hat{p}_{s-1}^{b_{s-1}^m}(\mathbf{y}|\theta_{s-1}^{b_{s-1}^m}, \mathcal{M}_i)$).

At the end of Algorithm 4, we can approximate the evidence $p(\mathbf{z}|\mathcal{M}_i)$ by

$$\hat{p}(\mathbf{z}|\mathcal{M}_i) \coloneqq \prod_{s=1}^{S} \frac{1}{M} \sum_{m=1}^{M} v_s^m.$$

The algorithm can also be used to infer parameters in the ith model. That is, a posterior expectation $\mathbb{E}[\varphi(\theta)]$, for $\theta \sim p(\theta|\mathbf{z}, \mathcal{M}_i)$ and test function φ, is approximated by $\sum_{m=1}^{M} V_S^m \varphi(\theta_S^m)$. Numerous extensions can make Algorithm 4 more efficient. The particular version of SMC sampler we use in our applications is given in Web Appendix C.

Other SMC samplers for model comparison can be found in Zhou et al. (2016). In addition, Chopin et al. (2013) proposed a related algorithm called *SMC²* which useful when one wishes to perform inference *sequentially* since it incorporates new data points as they arrive. However, as observed in Drovandi and McCutchan (2016), SMC^2 can become unstable when a newly arrived observation contradicts the existing information about the parameters. In such cases, the likelihood-tempering approach adopted here can lead to a smoother sequence of target distributions (Duan and Fulop, 2015) and hence more accurate estimates.

4.3 Improving SMC efficiency for integrated models

We are able to exploit the structure of integrated population models to enhance the efficiency of the SMC sampler for evidence approximation. Firstly, we employ DA approach from Subsection 3.2 to reduce the computational cost of the MCMC updates within the SMC sampler. Secondly, we propose to employ a likelihood-tempering approach which tempers the different parts of the likelihood separately. That is, for some $1 \leq S' < S$, the SMC sampler targets distributions of the form

$$\pi_s(\theta) \propto \begin{cases} p(\theta|\mathcal{M}_i)p(\mathbf{w}|\theta, \mathcal{M}_i)^{\alpha_s}, & \text{if } 0 \leq s \leq S', \\ p(\theta|\mathcal{M}_i)p(\mathbf{w}|\theta, \mathcal{M}_i)p(\mathbf{y}|\theta, \mathcal{M}_i)^{\beta_s}, & \text{if } S' < s \leq S, \end{cases}$$

where $0 = \alpha_0 < \alpha_1 < \ldots < \alpha_{S'} = 1$ and $0 < \beta_{S'+1} < \ldots < \beta_S = 1$. We note that the intractable count-data likelihood is again replaced by an unbiased estimate. The advantage of this refined tempering scheme is that the approximation of the count-data likelihood (obtained via the costly PF) is not needed in the first S' steps, so that S' can be taken to be large. Introducing the additional data likelihood first can be especially beneficial if these are informative about the parameters relative to the count data. This refined tempering strategy was crucial for obtaining reliable estimates in the herons example from Section 6; its efficiency gains are also illustrated in Subsection 5.4.

5 Example 1: Little owls

5.1 Parameters

The main model parameters – potentially specific to age group $a \in \{1, \mathrm{A}\}$ (1: juvenile, i.e. first-year, A: adult) and gender $g \in \{\mathrm{m}, \mathrm{f}\}$ (f: female, m: male) of the owls, and to time index $t \in \{1, \ldots, T\}$ – are

$\phi_{a,g,t}$**:** probability of an owl of gender g surviving to time $t+1$ if the owl is alive and in age group a at time t;

$p_{g,t+1}$**:** probability of observing a marked owl (in a capture-recapture setting) of gender g at time $t+1$ if alive at time $t+1$;

ρ_t**:** productivity rate governing the expected number of chicks produced per female at time t that survive to fledgling;

η_t**:** immigration rate governing the number of female immigrants at time $t+1$ per female of the population at time t.

5.2 Model specification

We consider the model defined by Schaub et al. (2006) and subsequently fitted in BUGS by Abadi et al. (2010b); see these papers for further information and ecological rationale.

5.2.1 Count-data model

The true population sizes for the juvenile and adult females, $\mathbf{x}_t = \{x_{1,t}, x_{\mathrm{A},t}\}$, evolve as

$$x_{1,t}|\mathbf{x}_{t-1}, \theta \sim \mathrm{Poisson}\big([x_{1,t-1} + x_{\mathrm{A},t-1}]\rho_{t-1}\phi_{1,\mathrm{f},t-1}/2\big), \quad x_{\mathrm{A},t} = sur_t + imm_t,$$

where $sur_t|\mathbf{x}_{t-1}, \theta \sim \mathrm{Binomial}(x_{1,t-1} + x_{\mathrm{A},t-1}, \phi_{\mathrm{A},\mathrm{f},t-1})$ is the number of female adults which survive from time $t-1$ to time t, and $imm_t|\mathbf{x}_{t-1}, \theta \sim \mathrm{Poisson}\big([x_{1,t-1} + x_{\mathrm{A},t-1}]\eta_{t-1}\big)$ is the number of female adults which immigrate in this period. We take the initial population sizes $x_{1,1}$, $x_{\mathrm{A},1}$ to be a-priori independently distributed according to a discrete uniform law on $\{0, 1, \ldots, 50\}$. The observation process is specified by $y_t|\mathbf{x}_t, \theta \sim \mathrm{Poisson}(x_{1,t}+x_{\mathrm{A},t})$.

5.2.2 Capture-recapture model

Capture-recapture data are available in the form of age-group and gender specific matrices $\mathbf{m} \coloneqq \{\mathbf{m}_{a,g} \colon a \in \{1, \mathrm{A}\}, g \in \{\mathrm{m}, \mathrm{f}\}\}$. The tth row, $\mathbf{m}_{a,g,t} \coloneqq \{m_{a,g,t,s} \colon 1 < s \leq T+1\}$, corresponds to the tth year of release ($t \in \{1, \ldots, T-1\}$). That is, $m_{a,g,t,s}$ is the number of individuals of gender g, last observed at age a at time t, that are recaptured at time s (if $t+1 \leq s \leq T$) or never recaptured again (if $s = T+1$). Note that $m_{a,g,t,s} = 0$ if $s \leq t$. For each year of release, we assume a multinomial distribution for the subsequent recaptures. The capture-recapture model specified as

$$\mathbf{m}_{a,g,t}|R_{a,g,t}, \theta \sim \mathrm{Multinomial}(R_{a,g,t}, \mathbf{q}_{a,g,t}).$$

Here, $R_{a,g,t}$ is number of owls in age group a and of gender g that observed (either an initial capture or, if $a = \mathrm{A}$, as a recapture) at time t and subsequently released. The multinomial cell probabilities $\mathbf{q}_{a,g,t} \coloneqq \{q_{a,g,t,s} \colon 1 < s \leq T+1\}$ are given by

$$q_{a,g,t,s} \coloneqq \begin{cases} 0, & \text{if } 1 < s \leq t, \\ \phi_{a,g,t} p_{g,s} \prod_{r=t+1}^{s-1} \phi_{\mathrm{A},g,r}(1 - p_{g,r}), & \text{if } t < s \leq T, \\ 1 - \sum_{r=1}^{T} q_{a,g,t,r}, & \text{if } s = T+1. \end{cases}$$

5.2.3 Fecundity model

Nest record data $\mathbf{n} := \{N_t, n_t \colon 1 \leq t \leq T\}$ are also available to provide information relating to the fecundity rate of little owls. Specifically, N_t denotes the number of breeding females recorded at time t and n_t the number of chicks produced that survive to leave the nest. Following Schaub et al. (2006) we specify $n_t | N_t, \theta \sim \text{Poisson}(N_t \rho_t)$. With this notation, the set of all additional data is $\mathbf{w} = \{\mathbf{m}, \mathbf{n}\}$.

5.3 Parametrisation and priors

There is additional covariate information about the abundance of voles – the primary source of prey for little owls – classified as low ($vole_t = 0$) or high ($vole_t = 1$), for each year of the study. Following Schaub et al. (2006); Abadi et al. (2010b), we parametrise $\log \eta_t = \delta_0 + \delta_1 vole_t$ and $\text{logit}\, p_{g,t+1} = \beta_1 \mathbb{I}\{g = \text{m}\} + \beta_{t+1}$ as well as

$$\text{logit}\, \phi_{a,g,t} = \alpha_0 + \alpha_1 \mathbb{I}\{g = \text{m}\} + \alpha_2 \mathbb{I}\{a = \text{A}\} + \alpha_3 year_t,$$

for $1 \leq t < T$, where the additional covariate: $year_t$, denotes the normalised year.

For simplicity, we assume that all components of θ have independent Normal$(0, 2)$ priors, except δ_0 for which we use a Normal$(-2, 2)$ prior because preliminary runs of the algorithm indicated that this parameter is typically very small. We avoided diffuse priors (a) to improve efficiency of the first steps of the SMC sampler and (b) to reduce the impact of the *Jeffreys–Lindley paradox* on the model comparison (i.e. very vague priors can unduly penalise a model; see Lindley, 1957). Other priors could have been employed but investigating the choice of priors in integrated population models is beyond the scope of the work.

5.4 Results

We end this section by demonstrating the gains in computational performance achievable for the PMCMC and SMC algorithms through the modifications proposed in Subsections 3.2 and 4.3. We also perform a model comparison to demonstrate the scientific

utility of our methodology.

Delayed acceptance. We first illustrate the performance gains obtained through the delayed-acceptance (DA) approach proposed in Subsection 3.2. For simplicity, we only report results for the model with productivity rate constant over time and immigration independent of the abundance of voles as this was one of the specifications which performed best in terms of model evidence. Figure 1 illustrates the utility of DA. It shows that even though DA decreases the acceptance rate (Christen and Fox, 2005), the computational savings due to only invoking PF for 'promising' parameter values more than compensate for this.

Refined tempering. In Table 1, we illustrate efficiency gains attainable through the refined likelihood tempering scheme (Section 4.3) over standard likelihood tempering (Section 4.2). For eight different models (specified in Web Appendix A), Table 1 displays

$$(\textit{efficiency gain}) = \frac{MSE \times (\textit{computation time})}{MSE \times (\textit{computation time})} \Big\}^{\text{standard tempering}}_{\text{refined tempering}}. \tag{6}$$

Here, *MSE* denotes the average mean-square error (MSE) of the estimate of the posterior mean based on 20 independent repeats of the SMC samplers (the average is taken over all components of the vector of model parameters); (*computation time*) represents the average computation time over the independent repeats. Since the true posterior means are intractable, we ran an MCMC algorithm using a large number $(10,000,000)$ of iterations for each model and treated the resulting posterior mean estimates as the true values.

Model comparison. Finally, we perform a model comparison to investigate the hypothesis from Abadi et al. (2010b) that little-owl immigration depends on the abundance of voles – their main prey. Figure 2 shows estimates of the evidence for the eight models specified in Web Appendix A in the case that the immigration rate may depend on the abundance of voles (i.e. $\delta_1 \neq 0$) and in the case that it is independent of vole abundance (i.e. $\delta_1 = 0$). The results indicate that the hypothesis is not supported by the data.

6 Example 2: Grey herons

6.1 Parameters

Younger herons typically having a lower survival probability than older adults. Following Besbeas et al. (2009), we therefore specify up to four age categories, indicated by the subscript $a \in \{1, \ldots, \mathrm{A}\}$; $a = 1$ represents first-years, $a = 2$ represents second-years and so on, while $a = \mathrm{A}$ represents all remaining adults. The main model parameters are

$\phi_{a,t}$**:** probability of a heron surviving until time $t + 1$ if alive and aged a at time t;

ρ_t**:** productivity rate w.r.t. the expected no. of females produced per female at time t;

λ_t**:** probability of recovering a dead heron in $[t, t + 1)$ if it died in that interval.

6.2 Model specification

For the specification of the integrated model, we follow Besbeas et al. (2009), allowing for some judicial modificatitions of the constituent state-space model.

6.2.1 Count-data model

We specify state-space model for the count data $\mathbf{y} = \{y_1, \ldots, y_T\}$. Let $x_{a,t}$, denote the true population sizes of herons in age group a at time t. These then evolve as

$$x_{a,t}|\mathbf{x}_{t-1}, \theta \sim \begin{cases} \text{Poisson}(\rho_{t-1}\phi_{1,t-1}\sum_{a=2}^{\mathrm{A}} x_{a,t-1}), & \text{if } a = 1, \\ \text{Binomial}(x_{a-1,t-1}, \phi_{a,t-1}), & \text{if } 1 < a < \mathrm{A}, \\ \text{Binomial}(x_{\mathrm{A}-1,t-1} + x_{\mathrm{A},t-1}, \phi_{\mathrm{A},t-1}), & \text{if } a = \mathrm{A}. \end{cases}$$

For simplicity, we assume that the distribution of each component of the initial state is a negative binomial distribution with probability $p = 1/100$ and size $n_0 = \mu_0 p/(1-p)$ for age groups $1 \leq a < \mathrm{A}$ and $n_1 = \mu_1 p/(1-p)$ for adults, respectively. We specify the means $\mu_0 = 5000/5$ and $\mu_1 = 5000-(\mathrm{A}-1)\mu_0$ in such a way that a-priori, $\mathbb{E}[\sum_{a=1}^{A} x_{a,1}|\theta] = 5000$.

Such a state-space model is typically approximated by a linear-Gaussian model in order to permit inference via the Kalman filter (Besbeas et al., 2002). However, the

assumption that $g_\theta(y_t|\mathbf{x}_t)$ is Gaussian is typically unrealistic, since it implies that the observation error is independent of scale and continuous. Instead, we consider a negative binomial observation process (with probability/size parametrisation), such that

$$y_t|\mathbf{x}_t, \theta \sim \text{Negative-Binomial}\left(\frac{\kappa}{1-\kappa}\sum_{a=2}^{A} x_{a,t}, \kappa\right),$$

for some $\kappa \in (0, 1)$. Note that this specification permits overdispersed observations since $\mathbb{E}[y_t|\mathbf{x}_t, \theta] = \sum_{a=2}^{A} x_{a,t} < \sum_{a=2}^{A} x_{a,t}/\kappa = \text{var}[y_t|\mathbf{x}_t, \theta]$.

6.2.2 Ring-recovery data model

Recall that count data are available from 1928 to 1998, i.e. for $T = 71$ time periods. In contrast, ring-recovery data are only available for individuals released between 1955 and 1997, i.e. released in time period $t \in \{t_1, \ldots, t_2\}$, where $t_1 = 28$ and $t_2 = 70$. These data are stored in a matrix $\mathbf{w}$ whose tth row is denoted $\mathbf{w}_t = \{w_{t,s} \colon t_1 + 1 \leq s \leq t_2 + 2\}$. Here, $w_{t,s}$ indicates the number of individuals released at time t which are subsequently recovered dead in the interval $(s-1, s]$; w_{t,t_2+2} corresponds to the number of individuals that are released at time t that are not seen again within the study.

For each year of release, we assume a multinomial distribution for the subsequent recoveries (see e.g. McCrea and Morgan, 2014, for further explanations of the ring-recovery model). Thus, the model for the rows of $\mathbf{w}$ is then specified as

$$\mathbf{w}_t|R_t, \theta \sim \text{Multinomial}(R_t, \mathbf{q}_t).$$

Here, R_t denotes the number of herons that are ringed as chicks and released in the tth time period. The multinomial cell probabilities $\mathbf{q}_t \coloneqq \{q_{t,s} \colon t_1 < s \leq t_2 + 2\}$ are given by

$$q_{t,s} \coloneqq \begin{cases} 0, & \text{if } t_1 < s \leq t, \\ (1 - \phi_{\min\{s-t,\text{A}\},s-1})\lambda_{s-1}\prod_{a=1}^{s-t-1}\phi_{\min\{a,\text{A}\},t+a-1}, & \text{if } t < s \leq t_2 + 1, \\ 1 - \sum_{s=t_1+1}^{t_2+1} q_{t,s}, & \text{if } s = t_2 + 2. \end{cases}$$

6.3 Parametrisation

6.3.1 Parameters common to all models

We consider additional covariate information to explain temporal variability. The recovery probabilities are assumed to be logistically regressed on the normalised covariate $time_t$ which represent the normalised (bird) year t:

$$\operatorname{logit} \lambda_t = \alpha_0 + \beta_0 time_t, \quad t = t_1, \ldots, t_2 - 1.$$

We specify the survival probabilities to be logistically regressed on the normalised covariate $fdays_t$ which represents the (normalised) number of days in (bird) year t on which the mean daily temperature fell below freezing in central England:

$$\operatorname{logit} \phi_{a,t} = \alpha_a + \beta_a fdays_t, \quad t = 1, \ldots, T - 1. \tag{7}$$

The free parameter in the negative-binomial observation equation is parametrised as $\kappa = \operatorname{logit}^{-1}(\omega) \in (0, 1)$ with $\omega \in \mathbb{R}$.

6.3.2 Models for the productivity rate

We specify a set of models – differing only in the productivity rate – for which we perform model comparison. The unknown parameters are $\theta = \{\omega, \alpha_0, \beta_0, \alpha_1, \ldots, \alpha_{\mathrm{A}}, \beta_1, \ldots, \beta_{\mathrm{A}}, \vartheta\}$, where ϑ represents the additional model parameters needed for one of the following models for the productivity rate.

Constant. We set $\log \rho_t = \psi$. Thus, $\vartheta = \{\psi\}$.

Regressed on frost days. We set $\log \rho_t = \gamma_0 + \gamma_1 fdays_{t-1}$. Thus, $\vartheta = \{\gamma_0, \gamma_1\}$.

Direct density dependence. Besbeas and Morgan (2012) investigated taking the log-productivity as a linear function of abundance, $\log \rho_t = \varepsilon_0 + \varepsilon_1 \tilde{y}_t$, where $\tilde{y}_t$ denotes the tth normalised observation. Thus, $\vartheta = \{\varepsilon_0, \varepsilon_1\}$.

Threshold dependence. Besbeas and Morgan (2012) also investigated models in which the productivity is a step function with K levels and hence $K - 1$ thresholds (K

itself may be unknown). The thresholds are a function of the observed rather than the true count to permit inference via the Kalman filter in Besbeas and Morgan (2012). The productivity rate is constant between change points and monotonically decreasing with increasing population size, i.e. for $K > 1$,

$$\rho_t = \begin{cases} \nu_1, & \text{if } y_t < \tau_1, \\ \nu_k, & \text{if } \tau_{k-1} \leq y_t < \tau_k \text{ for } 1 < k < K, \\ \nu_K, & \text{if } \tau_{K-1} \leq y_t, \end{cases}$$

where $\nu_1 > \nu_2 > \cdots > \nu_K$ and $\tau_1 < \tau_2 < \cdots < \tau_{K-1}$. Thus, it is assumed that larger population sizes induce lower productivity rates, e.g. due to an exhaustion of high quality breeding sites. To ensure these inequalities we set $\nu_K = \exp(\zeta_K)$ and

$$\nu_k = \textstyle\sum_{l=k}^{K} \exp(\zeta_l); \quad \tau_k = y_{\min} + (y_{\max} - y_{\min}) \dfrac{\sum_{l=1}^{k} \exp(\eta_l)}{\sum_{m=1}^{K} \exp(\eta_m)},$$

for $k \in \{1, \ldots, K-1\}$, where $y_{\min} = \min\{y_1, \ldots, y_T\}$ and $y_{\max} = \max\{y_1, \ldots, y_T\}$. In this case, $\vartheta = \{\zeta_k, \eta_k : 1 \leq k \leq K\}$.

Regime switching dynamics. To construct a more flexible model for the productivity rate, we extend the latent states $\mathbf{x}_t$ by including an additional (unobserved) regime indicator variable r_t which takes values in $\{1, \ldots, K\}$. Conditionally on r_{t-1}, the productivity rate ρ_{t-1} is then defined as $\rho_{t-1} = \nu_{r_{t-1}}$ where $\nu_1, \ldots, \nu_K$ are specified as in the threshold model, above. The evolution of the latent regime indicator r_t is assumed to be described by a Markov chain with transition equation

$$r_t | r_{t-1}, \theta \sim \text{Multinomial}(K, \mathbf{P}_{r_{t-1}}),$$

where $\mathbf{P}_k = (P_{k,1}, \ldots, P_{k,K})$ with $P_{k,l} = \exp(\varpi_{k,l}) / \sum_{m=1}^{K} \exp(\varpi_{k,m})$, for $1 \leq l \leq K$, is the kth row of the (K, K)-transition matrix for the regime indicator variable. In this case, $\vartheta = \{\zeta_k, \varpi_{k,l} \colon 1 \leq k, l \leq K\}$. This assumed Markovian structure allows the estimation of the productivity rate to borrow strength over time (which would not be possible in a temporal random-effect model).

Finally, we note that we also vary the number of levels, K, and the number of age

groups, A, so that the number of models to be compared is much larger than the five specifications for the productivity rate summarised above.

6.3.3 Prior specification

We assume that all the model parameters in θ are independent a-priori with Normal$(0, 1)$ priors, except that $\omega \sim$ Normal$(-2, 4)$. The motivation for this choice of priors is the same as in Subsection 5.3.

6.4 Results

Estimates of the evidence for the models can be found in Figure 3. The fit of the different models for the productivity rates is illustrated in Figure 4 below (see also Web Figure 2 in Web Appendix B). Due to the increased flexibility of the productivity rates, the regime-switching model leads to a smaller measurement error. In addition, the evidence for the regime-switching model is much higher than the other models in Figure 3.

Figure 3 supports the finding from Besbeas et al. (2009) that modelling the herons using four age groups is appropriate (though the results with three age groups are similar). Using only two age groups drastically reduces the model evidence across all specifications for the productivity rate. The results also support the findings from Besbeas and Morgan (2012) that the first three models (with productivity rate constant, regressed on the number of frost days, or density-dependent) do not explain the data well.

The posterior distribution of the productivity rate (under any of the models) must be interpreted with care. Indeed, note the sharp decline of the productivity rate in the years immediately preceding the severe winters of 1946–47 and 1962–63 in Figure 4b. This indicates that the linear model for the survival rates in (7) may not be flexible enough to accommodate the drop in the heron population in subsequent years.

We also implemented all of the above-mentioned models using a continuous (linear-Gaussian) approximation to the state-space model for the count data. The results (omitted here) are relatively similar to the results obtained for the original models. However, the regularising effect of the approximation artificially increased the evidence for all mod-

els by roughly the same amount, thus overestimating the model fit.

7 Conclusion

We have proposed methodology for Bayesian inference in integrated population models which have a state-space model for the noisily observed population sizes as one of their constituent parts. Utilising particle Markov chain Monte Carlo (PMCMC) techniques, our approach can be applied to fairly general models without the need for linear or Gaussian approximations which introduce a bias that is often difficult to quantify. Incorporating the approach into an SMC sampler also enables Bayesian model comparison, e.g. for the number of age groups. Finally, we have proposed extensions which exploit the integrated model structure to enhance the efficiency of our methodology.

We have demonstrated the methodology on two different applications: (1) little owls and (2) grey herons. For the owls, we have found no evidence in the data in favour of some complex model specifications proposed in the literature, e.g. for the dependence of immigration on the abundance of voles (Abadi et al., 2010b). For the herons, we have shown that existing models, including the elaborate threshold model for the productivity from Besbeas and Morgan (2012), do not explain the data well. As a remedy, we have proposed a novel regime-switching model and demonstrated that it is very strongly favoured over the competing models in terms of the Bayes factor. We note that the regime-switching model is motivated by statistical rather that ecological considerations. Its flexibility provides insights into the data and assists practitioners with the identification of factors affecting the productivity rate through the period of the study.

Our methodology is related to the SMC^2 algorithm from Drovandi and McCutchan (2016). However, even in low-dimensional settings (i.e. 3-4 unknown model parameters) Drovandi and McCutchan (2016) had to combine SMC^2 with a second importance-sampling algorithm to obtain evidence estimates accurate enough for model comparison in some examples (and, as pointed out by Drovandi and McCutchan (2016), this second scheme may not be applicable in higher dimensions). In contrast, in all applications

considered in this work, the evidence estimates provided by our methodology were accurate enough to directly identify the best-performing models despite the relatively large number of unknown model parameters (i.e. 6–58 for the owls; 8–31 for the herons).

The principle of using the evidence (and associated Bayes factors) for model comparison arises naturally in Bayesian statistics (Kass and Raftery, 1995) and eliminates the need for easier-to-compute but not-as-firmly-grounded alternatives such as the deviance information criterium (DIC) from Spiegelhalter et al. (2002) (see e.g. Pooley and Marion, 2018, for discussion and empirical comparisons) which was used e.g. in Abadi et al. (2010b). We note that overfitting is, in principle, not an issue when working with Bayes factors (for instance, a simple model is favoured in the little-owl example). Further posterior predictive checks could be carried out but this is beyond the scope of this work.

Future work in this area – potentially of great benefit to practitioners – could focus on making implementations of SMC samplers for model comparison available in easy-to-use software packages such as *Nimble* (de Valpine et al., 2017).

Acknowledgements

This work was supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1. A.B. and A.F. were funded by a Leverhulme Trust Prize.

Supplementary materials

Web Appendices and Figures, referenced in Sections 3–6, are available at [link to supplementary pdf file goes here]. All data and C++/R code necessary for reproducing the results can be found at https://github.com/AxelFinke/monte-carlo-rcpp.

References

Abadi, F., Gimenez, O., Arlettaz, R., and Schaub, M. (2010a). An assessment of integrated population models: Bias, accuracy, and violation of the assumption of independence. *Ecology*, 91(1):7–14.

Abadi, F., Gimenez, O., Ullrich, B., Arlettaz, R., and Schaub, M. (2010b). Estimation of immigration rate using integrated population models. *Journal of Applied Ecology*, 47(2):393–400.

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342. With discussion.

Bernardo, J. M. and Smith, A. F. M. (2009). *Bayesian Theory*. Wiley.

Besbeas, P., Borysiewicz, R. S., and Morgan, B. J. T. (2009). Completing the Ecological Jigsaw. In *Modeling Demographic Processes in Marked Populations*, pages 513–539. Springer.

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.

Besbeas, P. and Morgan, B. J. T. (2012). A threshold model for heron productivity. *Journal of Agricultural, Biological, and Environmental Statistics*, 17(1):128–141.

Breed, G., Costa, D., Jonsen, I., Robinson, P., and Mills-Flemming, J. (2012). State-space methods for more completely capturing behavioral dynamics from animal tracks. *Ecological Modelling*, 235:49–58.

Brooks, S. P., King, R., and Morgan, B. J. T. (2004). A Bayesian approach to combining animal abundance and demographic data. *Animal Biodiversity and Conservation*, 27(1):515–529.

Buckland, S. T., Newman, K. B., Fernandez, C., Thomas, L., and Harwood, J. (2007). Embedding population dynamics models in inference. *Statistical Science*, 22(1):44–58.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1).

Chopin, N. (2002). A sequential particle filter method for static models. *Biometrika*, 89(3):539–552.

Chopin, N., Jacob, P. E., and Papaspiliopoulos, O. (2013). SMC^2: An efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(3):397–426.

Christen, J. A. and Fox, C. (2005). Markov chain Monte Carlo using an approximation. *Journal of Computational and Graphical Statistics*, 14(4):795–810.

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., and Bodik, R. (2017). Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26(2):403–413.

Del Moral, P. (1996). Nonlinear filtering: Interacting particle solution. *Markov Processes and Related Fields*, 2(4):555–580.

Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B*, 68(3):411–436.

Doucet, A. and Johansen, A. M. (2011). A tutorial on particle filtering and smoothing: Fifteen years later. In Crisan, D. and Rozovskii, B., editors, *The Oxford Handbook of Nonlinear Filtering*, Oxford Handbooks, chapter 24, pages 656–704. Oxford University Press.

Drovandi, C. C. and McCutchan, R. A. (2016). Alive SMC^2: Bayesian model selection for low-count time series models with intractable likelihoods. *Biometrics*, 72(2):344–353.

Duan, J.-C. and Fulop, A. (2015). Density-tempered marginalized sequential Monte Carlo samplers. *Journal of Business & Economic Statistics*, 33(2):192–202.

Dupuis, J. A. (1995). Bayesian estimation of movement and survival probabilities from capture-recapture data. *Biometrika*, 82(4):761–772.

Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. (1994). A language and program for complex Bayesian modelling. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 43(1):169–177.

Golightly, A., Henderson, D. A., and Sherlock, C. (2015). Delayed acceptance particle MCMC for exact inference in stochastic kinetic models. *Statistics and Computing*, 25(5):1039–1055.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45.

Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

King, R. (2011). Statistical Ecology. In Brooks, S., Gelman, A., Jones, G., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*, chapter 17, pages 410–447. CRC Press.

King, R. (2012). A review of Bayesian state-space modelling of capture-recapture-recovery data. *Interface Focus*, 2:190–204.

King, R. (2014). Statistical ecology. *Annual Review of Statistics and its Application*, 1(1):401–426.

King, R., Brooks, S., Mazzetta, C., Freeman, S., and Morgan, B. (2008). Identifying and diagnosing population declines: A Bayesian assessment of lapwings in the UK. *Journal of Royal Statistical Society: Series C*, 57(5):609–632.

Knape, J. and de Valpine, P. (2012). Fitting complex population models by combining particle filters with Markov chain Monte Carlo. *Ecology*, 93(2):256–263.

Lindley, D. V. (1957). A statistical paradox. *Biometrika*, 44(1/2):187–192.

McClintock, B. T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B. J., and Morales, J. M. (2012). A general discrete-time modeling framework for animal movement using multi-state random walks. *Ecological Monographs*, 82(3):335–349.

McCrea, R. S. and Morgan, B. J. T. (2014). *Analysis of Capture-Recapture Data*. CRC Press.

Millar, R. B. and Meyer, R. (2000). Non-linear state space modelling of fisheries biomass dynamics by using Metropolis-Hastings within-Gibbs sampling. *Journal of the Royal Statistical Society: Series C*, 49(3):327–342.

Morales, J., Haydon, D., Frair, J., Holsiner, K., and Fryxell, J. (2004). Extracting more out of relocation data: Building movement models as mixtures of random walks. *Ecology*, 85(9):2436–2445.

Newman, K. B. (1998). State-space modelling of animal movement and mortality with application to salmon. *Biometrics*, 54:1290–1314.

Newman, K. B., Buckland, S. T., Morgan, B. J. T., King, R., Borchers, D. L., Cole, D., Besbeas, P. T., Gimenez, O., and Thomas, L. (2014). *Modelling Population Dynamics: Model Formulation, Fitting and Assessment using State-space Methods*. Springer.

Nishimura, A., Dunson, D., and Lu, J. (2017). Discontinuous Hamiltonian Monte Carlo for models with discrete parameters and discontinuous likelihoods. *ArXiv e-prints*, 1705.08510.

Parslow, J., Cressie, N., Campbell, E. P., Jones, E., and Murray, L. (2013). Bayesian learning and predictability in a stochastic nonlinear dynamical model. *Ecological Applications*, 23(4):679–698.

Peters, G. W., Hosack, G. R., and Hayes, K. R. (2010). Ecological non-linear state space model selection via adaptive particle Markov chain Monte Carlo (AdPMCMC). *ArXiv e-prints*, 1005.2238.

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

Pooley, C. and Marion, G. (2018). Bayesian model evidence as a practical alternative to deviance information criterion. *Royal Society Open Science*, 5(3):171519.

Royle, J. A. (2008). Modeling individual effects in the Cormack-Jolly-Seber model: A state-space formulation. *Biometrics*, 64(2):364–370.

Schaub, M., Ullrich, B., Knötzsch, G., Albrecht, P., and Meisser, C. (2006). Local population dynamics and the impact of scale and isolation: A study on different little owl populations. *Oikos*, 115(3):389–400.

Sherlock, C., Thiery, A., and Golightly, A. (2015). Efficiency of delayed-acceptance random walk Metropolis algorithms. *ArXiv e-prints*, 1506.08155.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Zhou, Y., Johansen, A. M., and Aston, J. A. D. (2016). Towards automatic model comparison: An adaptive sequential Monte Carlo approach. *Journal of Computational and Graphical Statistics*, 25(3):701–726.

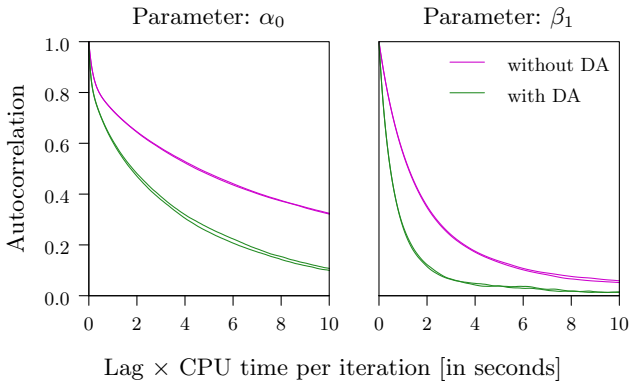


Figure 1. Autocorrelation (rescaled by computation time) of the estimates of the parameters α_0 and β_1 in the little-owls model (with the productivity rates assumed to be constant, i.e. $\rho_1 = \ldots = \rho_T$) and immigration independent of the abundance of voles, i.e. $\delta_1 = 0$. The results are based on two independent repeats (each comprised of 10^7 iterations) of the MCMC algorithms with and without delayed-acceptance.

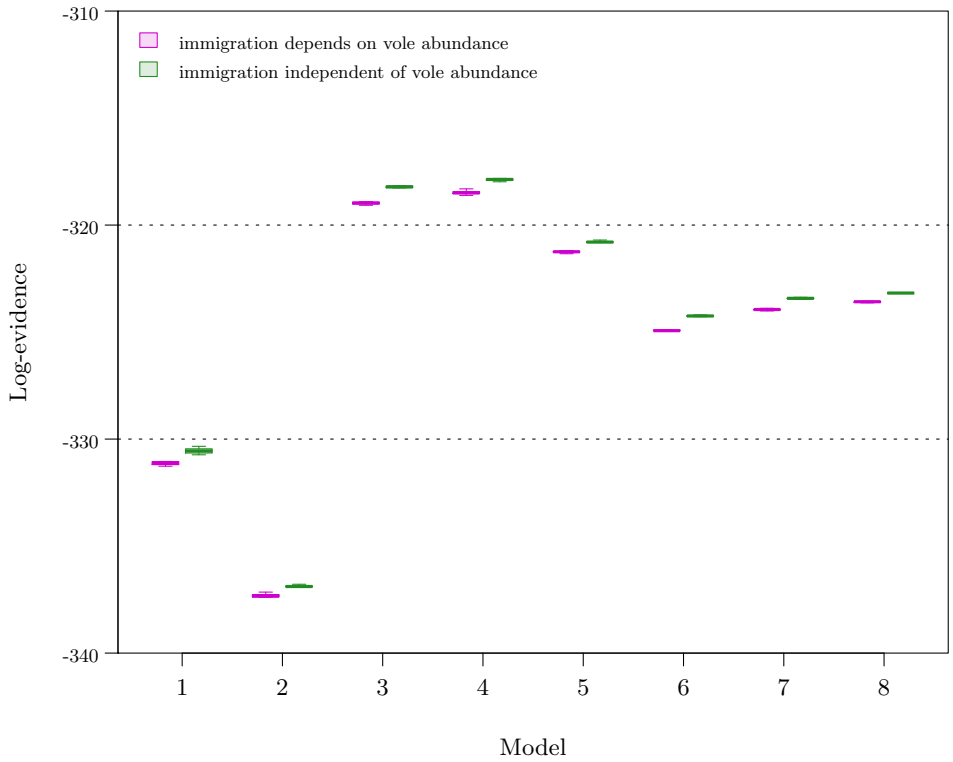


Figure 2. Logarithm of the estimates of the evidence for the eight models for the little owls with or without dependence of the immigration rate on the abundance of voles. The results were obtained from 20 independent runs of the adaptive SMC sampler using $10,000$ particles; the particle filters used to approximate the marginal likelihoods use $1,000$ particles. The average computation time for each SMC sampler was around 9–18 hours on a single core. We stress such a relatively large number of particles was only used to gain accurate evidence estimates in the more complex models (in terms of the number of parameters), i.e. in Models 1–5. For the smaller models, i.e. Models 6–8, quite similar results could have been obtained in 30 minutes by using only 500 particles.

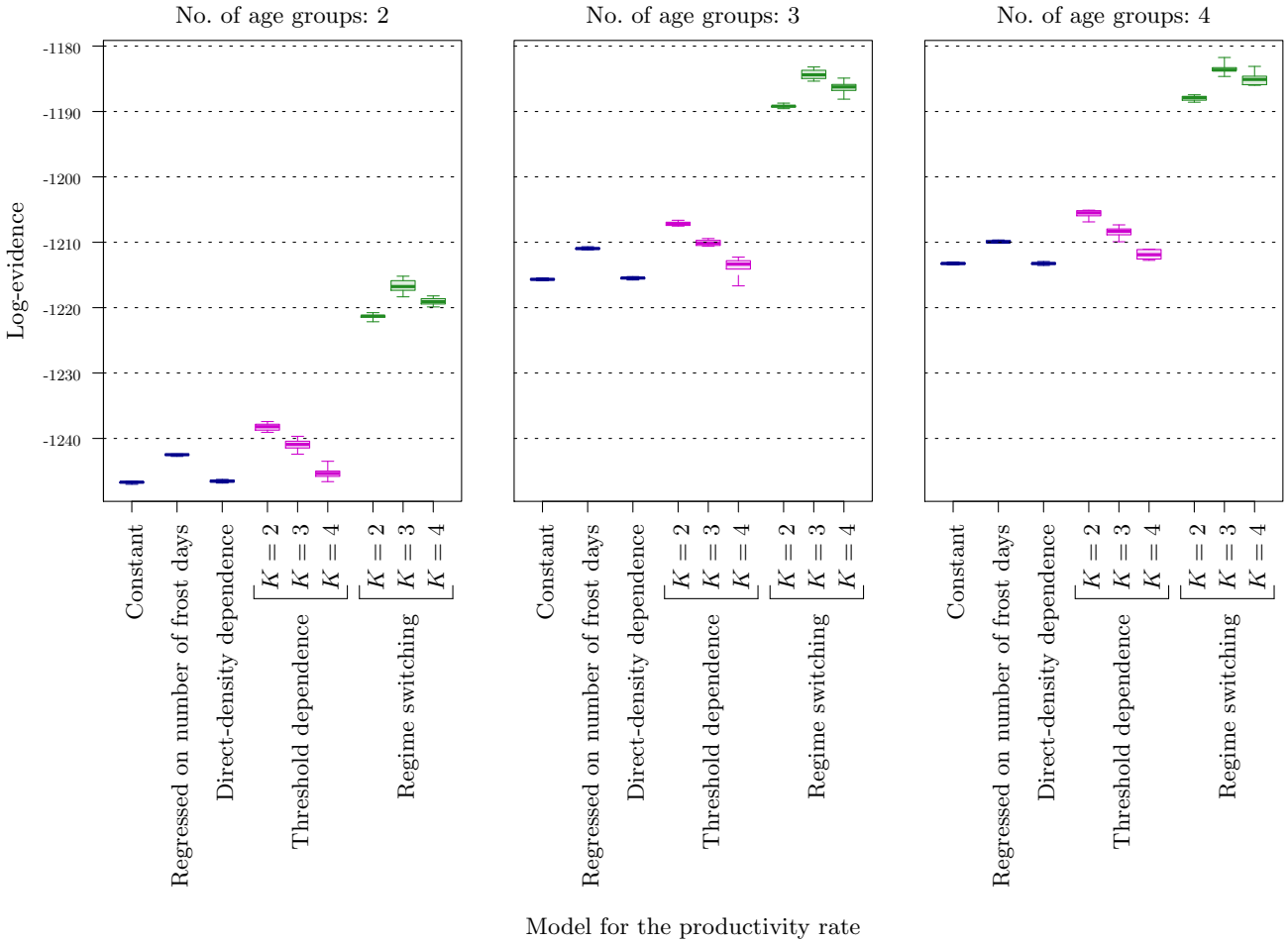
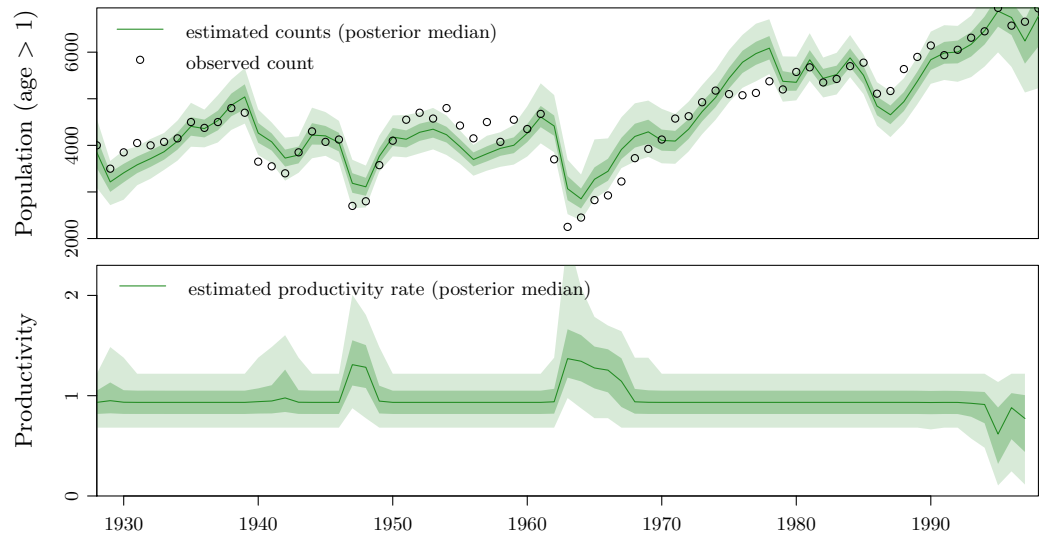
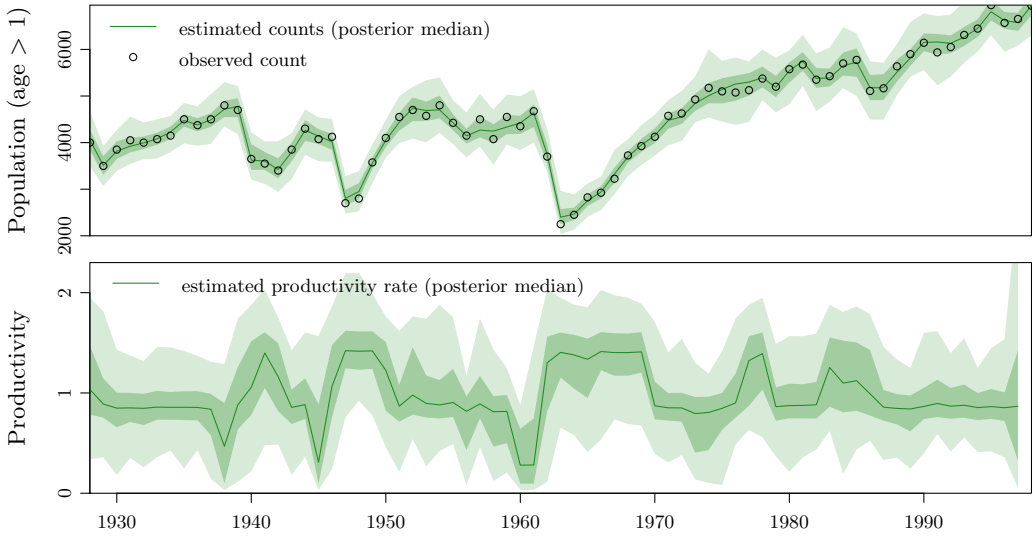


Figure 3. Logarithm of estimates of the evidence for different models for the grey herons. Shown are results for the different models for the productivity rate and different numbers of distinctly modelled age categories (A). For the threshold and regime-switching models, we also investigate different values for the number of thresholds/regimes (K). Obtained from 10 independent runs of the adaptive SMC sampler using $1,000$ particles; the PFs used to approximate the count-data likelihood employed $4,000$ particles. The average computation time was 42–61 hours for the threshold models, 32–45 hours for the regime-switching models and 29–48 for the remaining models, the lower numbers corresponding to A $= 2$ age categories and the higher numbers to A $= 4$ age categories.



a. Threshold dependence ($K = 4$ levels, i.e. 3 thresholds).



b. Regime switching ($K = 4$ regimes).

Figure 4. Marginal posterior distributions of the estimated heron counts (top rows) and productivity rates (bottom rows) for the threshold model from Besbeas and Morgan (2012) and the novel regime-switching model (results for other models are shown in Web Appendix B) with A = 4 distinct age categories. The shaded areas represent, respectively, the 90 % quantile and range of all encountered realisations. The shown results display the average over 10 independent repeats of the adaptive SMC sampler (each using $1,000$ particles). The PFs used to approximate the count-data likelihood use $4,000$ particles.

Table 1. Average efficiency gain (as defined in Equation (6)) of the refined likelihood tempering scheme (see Section 4.3 of the main manuscript) over standard likelihood tempering for different numbers of particles (M). To simplify the presentation, we only show results for each of the eight models in the case that $\delta_1 \neq 0$, i.e. we allow for dependence of immigration on the abundance of voles.

Model	1	2	3	4	5	6	7	8
efficiency gain ($M = 1,000$)	14.0	4.7	2.8	2.3	2.9	0.9	0.9	1.2
efficiency gain ($M = 10,000$)	19.1	4.5	2.3	2.3	2.5	0.8	1.2	0.6