



# Kent Academic Repository

**Besbeas, Panagiotis and Morgan, Byron J. T. (2019) *Exact inference for integrated population modelling*. *Biometrics*, 75 (2). pp. 475-484. ISSN 0006-341X.**

## Downloaded from

<https://kar.kent.ac.uk/71374/> The University of Kent's Academic Repository KAR

## The version of record is available from

<https://doi.org/10.1111/biom.13045>

## This document version

Author's Accepted Manuscript

## DOI for this version

## Licence for this version

UNSPECIFIED

## Additional information

## Versions of research works

### Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

### Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal*, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

## Enquiries

If you have questions about this document contact [ResearchSupport@kent.ac.uk](mailto:ResearchSupport@kent.ac.uk). Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

**Exact inference for integrated population modelling<sup>†</sup>**

**P. Besbeas<sup>1,2\*</sup>, B.J.T. Morgan<sup>2\*\*</sup>**

<sup>1</sup> Department of Statistics, Athens University of Business and Economics, 10434 Athens, Greece;

<sup>2</sup> National Centre for Statistical Ecology, School of Mathematics, Statistics and Actuarial Science,  
University of Kent, Canterbury, Kent CT2 7FS, England.

\**email*: P.T.Besbeas@kent.ac.uk

\*\**email*: B.J.T.Morgan@kent.ac.uk

This paper has been submitted for consideration for publication in *Biometrics*

†This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1111/biom.13045]

**Additional Supporting Information may be found in the online version of this article.**

**Received 27 February 2018; Revised 14 November 2018; Accepted 6 December 2018**

**Biometrics**

**This article is protected by copyright. All rights reserved**

**DOI 10.1111/biom.13045**

## Summary:

Integrated population modelling is widely used in statistical ecology. It allows data from population time series and independent surveys to be analysed simultaneously. In classical analysis the time-series likelihood component can be conveniently approximated using Kalman filter methodology. However, the natural way to model systems which have a discrete state space is to use hidden Markov models (HMMs). The proposed method avoids the Kalman filter approximations and Monte Carlo simulations. Subject to possible numerical sensitivity analysis, it is exact, flexible, and allows the use of standard techniques of classical inference. We apply the approach to data on Little owls, where the model is shown to require a one-dimensional state space, and Northern lapwings, with a two-dimensional state space. In the former example the method identifies a parameter redundancy which changes the perception of the data needed to estimate immigration in integrated population modelling. The latter example may be analysed using either first- or second-order HMMs, describing numbers of one-year olds and adults or adults only, respectively. The use of first-order chains is found to be more efficient, mainly due to the smaller number of one-year olds than adults in this application. For the lapwing modelling it is necessary to group the states in order to reduce the large dimension of the state space. Results check with Bayesian and Kalman filter analyses, and avenues for future research are identified. This article is protected by copyright. All rights reserved

Key words: Hidden Markov models; Little owls; migration; Northern lapwings; parameter redundancy; state-space models

## 1. Introduction

Integrated population modelling (IPM) is the state-of-the-art approach for estimating parameters of population dynamics when independent data sets are available at the population and individual levels on members of the same wild animal population. These data sets typically relate to animal survival, productivity and abundance, in the last case through time series of counts. The models can be fitted by maximum-likelihood (Besbeas et al., 2002; deValpine, 2012) or computational Bayesian methods (Brooks et al., 2004; Kéry and Schaub, 2012, Chapter 11). Important demographic parameters for which there is no direct survey information might be estimated using IPM: this was productivity in the case of Besbeas et al. (2002) and immigration in the case of Abadi et al. (2010). Literature surveys of IPM are provided by Schaub and Abadi (2011), and in fisheries science, by Maunder and Punt (2013). For recent research in IPM see for example Besbeas and Morgan (2017), Finke et al. (2019) and Lahoz-Monfort et al. (2017).

The aim of this paper is to show how to utilise efficient hidden Markov model (HMM) methodology to provide the time-series likelihood, which is typically central to IPM. This then allows exact IPM using maximum likelihood, and provides useful tools from classical inference, including model comparison and goodness-of-fit. Bayesian analysis is also exact, but requires Markov chain Monte Carlo. The new approach is flexible, avoids making the assumptions involved in using the Kalman filter to approximate the likelihood for population time-series data, and is simpler than the alternative approaches of deValpine (2012) and Knappe et al. (2011), the latter of which focusses on modelling time series of population counts alone.

In Section 2 we describe the two case studies of the paper. In Section 3 we present models for data from studies of capture-recapture, ring-recovery and productivity. We describe the main current methods that are used to model population time-series, and introduce the

HMM approach. We also explain how component likelihoods for independent data sets are combined to form a single, integrated likelihood. Section 4 illustrates the HMM method of this paper on the two data sets. Comparisons are made with the results of Bayesian analysis and using the Kalman filter. Section 5 outlines the potential of the HMM approach and new avenues for research.

## 2. Data

### 2.1 *Little owl, *Athene noctua**

The data are available from the supplementary material for Abadi et al. (2010). They describe data on the Little owl, obtained from 1978 to 2003 from birds nesting in nest boxes in Göppingen, providing recapture information on survival, stratified by age and sex, as well as data on productivity and on population size. The primary prey of Little owls is voles, and annual spring vole abundance is described by means of a binary covariate, indicating either high or low abundance.

### 2.2 *Northern lapwing, *Vanellus vanellus**

Two data sets provide information on survival and counts for the Northern lapwing; there is no sex information and age is known for survival. The count data were collected from 1965 – 1998, and are illustrated in Besbeas et al. (2002). They are obtained from 447 sites surveyed under the Common Birds Census (CBC) of the British Trust for Ornithology (BTO), and may be regarded as providing information on the total population of lapwings for those 447 sites. Birds were ringed as nestlings between 1963–1997 and ring-recovery data were obtained from the reporting of dead birds. In addition a covariate provides the number of days between April in year  $t$  and March in year  $t+1$  that the temperature at a central England location was below freezing, which is used to model survival. The complete data, including a transformed

67 version of the covariate, are embedded in the WinBUGS code provided by Brooks et al.  
68 (2004).

### 69 3. Component and integrated modelling

70 Throughout we use boldface to indicate generic parameters which may involve several coef-  
71 ficients due to variation over time and/or by age.

#### 72 3.1 *Survival*

73 Ring-recovery and recapture data each result in multinomial distributions for the numbers  
74 of marked animals encountered in successive years following marking, recorded as dead in  
75 the case of recovery, and alive in the case of recapture. For either type of data the likelihood  
76 is then the product of multinomial probabilities, with one multinomial for each year of  
77 the study, parameterised in terms of annual survival probabilities. Note that we adopt the  
78 standard convention that, in modelling recovery data  $\mathbf{S}$  denotes annual survival probabilities,  
79 and in modelling recapture data  $\phi$  denotes apparent survival probabilities, with elements that  
80 are products of survival and retention probabilities. Models are completed with appropriate  
81 nuisance probability parameters for recovery,  $\boldsymbol{\lambda}$ , or recapture  $\mathbf{p}$ , as appropriate; see McCrea  
82 and Morgan (2014, Chapter 4).

83 We denote the likelihood for capture-recapture data as  $L_C(\phi, \mathbf{p}; \mathbf{m})$ , in which we use  $\mathbf{m}$   
84 to denote the matrix of numbers of recaptures, commonly called the m-array (McCrea and  
85 Morgan, 2014, p.69). The corresponding notation adopted for the likelihood for recovery data  
86 is  $L_R(\mathbf{S}, \boldsymbol{\lambda}; \mathbf{d})$ , where  $\mathbf{d}$  is the matrix of numbers of recoveries. For both matrices, each row  
87 contains the recorded numbers for the year of release of marked individuals corresponding  
88 to that row, and the columns indicate the year of recovery or recapture, respectively. The  
89 vector  $\mathbf{R}$  provides the annual totals of marked birds released. For illustration we provide  
90 below formulations for when there are fully time-dependent parameters, for annual studies

of length  $T$  years. There will be straightforward extensions to incorporate degrees of age-dependence in survival in both the case studies.

**3.1.1 Cormack-Jolly-Seber (CJS) model.** The basic CJS model has time-dependent parameters for apparent survival and recapture probability; see McCrea and Morgan (2014, p.70).

We define apparent survival  $\phi_i$ , for animals alive at time  $t_i$  which remain in the study area until time  $t_{i+1}$  and define  $p_j$  as the probability an individual which is alive at occasion  $t_j$  is recaptured at that time. The probability associated with the  $(i, j)$  cell of the m-array is then given by:

$$\nu_{ij} = \left\{ \prod_{k=i}^{j-1} \phi_k \prod_{\ell=i+1}^{j-1} (1 - p_\ell) \right\} p_j \quad \text{for } 1 \leq i < j \leq T,$$

and we define  $\chi_i = 1 - \sum_{j=i+1}^T \nu_{ij} = 1 - \phi_i \{1 - (1 - p_{i+1})\chi_{i+1}\}$ , for  $1 \leq i < T$ , and  $\chi_T = 1$ .

The product-multinomial likelihood is then given by

$$L_C(\boldsymbol{\phi}, \mathbf{p}; \mathbf{m}) \propto \prod_{i=1}^{T-1} \prod_{j=i+1}^T \nu_{ij}^{m_{ij}} \times \chi_i^{R_i - \sum_{j=i+1}^T m_{ij}}, \quad (1)$$

where  $R_i$  is the number of marked animals released at time  $t_i$ . The CJS model is parameter redundant as not all of the apparent survival and capture probabilities can be estimated: parameters  $\phi_{T-1}$  and  $p_T$  are confounded and only their product can be estimated. However all the other probabilities can in principle be estimated, and explicit expressions exist for maximum-likelihood estimates (McCrea and Morgan, 2014, pp 70–71). We build on this model in Section 4.1.

**3.1.2 Ring-recovery model.** We illustrate the likelihood with time-dependent survival probabilities  $\{S_i\}$  and probabilities  $\{\lambda_j\}$  for the reporting of dead animals. We assume that  $d_{ij}$  individuals from the  $i^{\text{th}}$  cohort of marked individuals are reported dead at time  $t_j$ . Making use of the assumption of independence of individuals between cohorts, the data can be modelled

by a product of multinomials, as above, and the likelihood is now given by

$$L_R(\mathbf{S}, \boldsymbol{\lambda}; \mathbf{d}) \propto \prod_{i=1}^{T-1} \prod_{j=i+1}^T \delta_{ij}^{d_{ij}} \times \epsilon_i^{R_i - \sum_{j=i+1}^T d_{ij}}, \quad (2)$$

where  $R_i$  denotes the number of marked individuals released at time  $t_i$ ,

$$\delta_{ij} = \begin{cases} (1 - S_i)\lambda_i & i = j - 1 \\ \prod_{k=1}^{j-2} S_k(1 - S_{j-1})\lambda_{j-1} & i < j - 1, \end{cases}$$

for  $1 \leq i < j \leq T$ , and  $\epsilon_i = 1 - \sum_{j=i+1}^T \delta_{ij}$ , for  $1 \leq i < T$ .

### 3.2 Productivity

Abadi et al. (2010) adopt yearly estimation of time-dependent model fecundities, assumed to have Poisson distributions,  $\{J_t\} \sim \text{Pois}(V_t r_t)$ , where  $V_t$  is the known number of reproducing females,  $r_t$  is the common individual productivity, and  $J_t$  is the number of total recorded offspring, all in year  $t$ . Assuming independence across years, we can write the likelihood for the productivity data alone as

$$L_P(\mathbf{r}; \mathbf{j}, \mathbf{V}) \propto \prod_{t=1}^T e^{-V_t r_t} (V_t r_t)^{j_t},$$

and we use this likelihood in the Little owl data analysis. Thus taken in isolation, fitting the productivity data of year  $t$ , with  $J_t = j_t$ , the model for productivity results in the maximum-likelihood estimates,  $\hat{r}_t = j_t/V_t$ , with estimated standard errors  $\sqrt{j_t/V_t}$ .

### 3.3 Population counts

Models for population time-series data are state-space models, taking discrete values, typically at integer times, which is true of the two case studies. Virtually all existing IPM has included state-space modelling of the population time-series data, which we take as annual. We denote the unobserved state vector at time  $t$  by  $\mathbf{N}_t = (N_{1,t}, \dots, N_{K,t})'$ , for  $t = 1, \dots, T$ , where  $N_{j,t}$  is the number of individuals in state  $1 \leq j \leq K$  at time  $t$ , and the annual population counts,  $\{\mathbf{y}_t\}$ , form an  $M$ -variate time series, for  $M \geq 1$ . The general formulation for state-space models links the state and observation processes as follows (Newman et al.,

2014, p. 43)

$$\mathbf{N}_1 \sim g_1(\mathbf{n}|\boldsymbol{\theta}), \quad (3)$$

$$\mathbf{N}_{t+1}|\mathbf{N}_t \sim g_t(\mathbf{n}|\mathbf{N}_t, \boldsymbol{\theta}), \quad \text{for } t \geq 1, \quad (4)$$

$$\mathbf{y}_t|\mathbf{N}_t \sim f_t(\mathbf{y}|\mathbf{N}_t, \boldsymbol{\psi}), \quad (5)$$

108 for an initial state distribution,  $g_1$ , a state distribution at time  $t$ ,  $g_t$ , and an observation  
 109 distribution  $f_t$ , where  $\boldsymbol{\theta}$  denotes model parameters for the state process and  $\boldsymbol{\psi}$  are parameters  
 110 for the observation process. The state distribution  $g_t$  can be extended to greater than first-  
 111 order dependence. We just consider linear models, although the approach of the paper is  
 112 general, except when the Kalman filter is used; see Besbeas and Morgan (2018).

113 We write the likelihood for the time-series data when survival estimation is based upon  
 114 capture-recapture data as  $L_T(\boldsymbol{\phi}, \mathbf{r}, \boldsymbol{\sigma}, \mathbf{N}_1; \mathbf{y}_t)$ , if variance parameters  $\boldsymbol{\sigma}$  for the observation  
 115 equation and  $\mathbf{N}_1$  are included in the model, and similarly for recovery data.

3.3.1 *Kalman filter.* Besbeas et al. (2002) provide a convenient approximation to  $L_T$  for  
 state-space population dynamics models based on the Kalman filter. Appropriate discrete  
 state distributions such as Poisson and binomial, are suitably approximated by normal  
 distributions, and the observation distributions are also taken as normal. Thus corresponding  
 to Equations 4 and 5, we have the multivariate normal distributions,

$$\mathbf{N}_{t+1}|\mathbf{N}_t \sim N(\boldsymbol{\Lambda}_t \mathbf{N}_t, \boldsymbol{\Omega}) \quad \text{for } t \geq 1, \quad (6)$$

$$\mathbf{y}_t|\mathbf{N}_t \sim N(\mathbf{Z}_t \mathbf{N}_t, \boldsymbol{\Sigma}) \quad \text{for } t \geq 1, \quad (7)$$

116 where  $\boldsymbol{\Lambda}_t$  is a  $K \times K$  Leslie matrix,  $\mathbf{Z}_t$  is an appropriate  $M \times K$  matrix and  $\boldsymbol{\Omega}$  and  $\boldsymbol{\Sigma}$  are  
 117 dispersion matrices.

118 In addition random variables appearing in the variance terms in the state Equation 6  
 119 are approximated by their expectations. The likelihood is easily formed, the method is

120 fast, performs well and is robust with respect to departures from the assumptions and  
 121 approximations made, even for small population sizes; see Brooks et al. (2004).

122 **3.3.2 Hidden Markov models.** Discrete time-series data can in principle be fitted exactly  
 123 by classical inference using the efficient machinery of HMMs, without the need of the  
 124 approximations used in Kalman filter analysis; see Cowen et al. (2017), King (2012), King  
 125 (2014), and Zucchini et al. (2016) for general introductions and applications of HMMs.  
 126 Exact analysis also facilitates extensions such as incorporation of density dependence in  $\mathbf{\Lambda}_t$ ;  
 127 cf Besbeas and Morgan (2012). The approach involves setting an upper bound,  $N_{max}$ , for  
 128 each state variable in the model, resulting in a finite  $N_{max}^K$ -state Markov chain; we generalise  
 129 this notation later. As the state vector for the approach adopted describes the number of  
 130 individuals in a population, potentially also stratified by age, then the dimension of the state  
 131 space may become large, and we shall discuss alternative ways of dealing with this feature.

In general, a HMM likelihood  $L_T$ , can be written as a product of the initial distribution vector  $\boldsymbol{\delta}$ , corresponding to  $g_1(\mathbf{n}|\boldsymbol{\theta})$  of Equation 3, the appropriate, year-dependent, transition probability matrices  $\{\boldsymbol{\Gamma}_t\}$ , corresponding to  $g_t(\mathbf{n}|\mathbf{N}_t, \boldsymbol{\theta})$  of Equation 4 which describe the state transitions in the latent process, and the state-dependent probability matrices  $\{\mathbf{P}(\mathbf{y}_t)\}$  for each year  $t$ , for the observation process, corresponding to  $f_t(\mathbf{y}|\mathbf{N}_t, \boldsymbol{\psi})$  of Equation 5. We can then write

$$L_T = \boldsymbol{\delta} \mathbf{P}(\mathbf{y}_1) \boldsymbol{\Gamma}_1 \cdots \boldsymbol{\Gamma}_{T-1} \mathbf{P}(\mathbf{y}_T) \mathbf{1}', \quad (8)$$

132 where  $\mathbf{1}$  denotes the unit row vector, which is the standard forward probability formulation  
 133 for HMM likelihoods (Zucchini et al., 2016, p. 37).

134 **3.3.3 Bayesian inference.** The Bayesian approach uses MCMC and also does not need  
 135 to make the assumptions of the classical analysis using the Kalman filter; see Brooks et al.,

2004 and Kéry and Schaub, 2012, Section 11.2, who describe a state-space model for the  
time-series likelihood.

### 3.4 Integrated population modelling

Under the assumption that the data from the different surveys are independent, the likelihood for integrated modelling,  $L_I$ , is given as the product of the corresponding component likelihoods. Then for capture-recapture, for example, and the models of Sections 3.1.1, 3.2 and 3.3 we obtain

$$L_I(\mathbf{r}, \phi, \sigma, \mathbf{p}, \mathbf{N}_1; \mathbf{j}, \mathbf{V}, \mathbf{y}_t, \mathbf{m}) = L_P(\mathbf{r}; \mathbf{j}, \mathbf{V})L_T(\phi, \mathbf{r}, \sigma, \mathbf{N}_1; \mathbf{y}_t)L_C(\phi, \mathbf{p}; \mathbf{m}), \quad (9)$$

with a similar equation for recovery data when productivity data are available. The expressions when productivity data are absent are obvious. In Equation 9, the likelihood,  $L_I$ , for the time-series data is pivotal, as it links the likelihood components together through common parameters. In classical inference maximisation takes place of  $L_I$ , with respect to all of the model parameters, while for Bayesian inference the posterior distribution is the product of  $L_I$  and the appropriate joint prior distribution. Note that  $\mathbf{N}_1$  and/or  $\sigma$  may not form part of the parameter set, as explained below.

## 4. Specific models and results for the case studies

### 4.1 Estimating immigration of Little owls

Abadi et al. (2010) use Bayesian inference to fit an IPM which integrates models for data on capture-recapture, productivity and also population count data. The model for (apparent) survival and capture probability assumed by Abadi et al. (2010) is the CJS model of Section 3.1.1 extended to include specific sex and age effects. There are two age classes for apparent survival: for birds aged one year, and for all older birds (taken not to vary with age); there is logistic-linear regression on year and additive parameters to distinguish between age and sex. The recapture probability has a different value for each year, and an additive parameter on

the logistic scale to distinguish sex. Productivity also has a different value for each year. We parameterise the model using standard logistic and logarithmic transformations as follows:

$$\text{logit}(\phi_{f,1,t}) = \beta_0 + \beta_1 t; \quad \text{logit}(\phi_{f,a,t}) = \beta_0 + \theta + \beta_1 t$$

$$\text{logit}(\phi_{m,1,t}) = \beta_0 + \delta + \beta_1 t; \quad \text{logit}(\phi_{m,a,t}) = \beta_0 + \theta + \delta + \beta_1 t$$

$$\text{logit}(p_{f,t}) = \kappa_t; \quad \text{logit}(p_{m,t}) = \zeta + \kappa_t; \quad \log(r_t) = \xi_t.$$

Here  $t$  indicates year,  $\phi_{f,1,t}$  ( $\phi_{m,1,t}$ ) is the survival probability of female (male) birds in their first year of life at time  $t$ ;  $\phi_{f,a,t}$  ( $\phi_{m,a,t}$ ) is the survival probability of older female (male) birds and  $p_{f,t}$  ( $p_{m,t}$ ) is the recapture probability of female (male) birds at time  $t$ . The complete parameter set for modelling survival is:  $\{\beta_0, \beta_1, \theta, \delta, \{\kappa_t\}, \zeta\}$ . Abadi et al. (2010) and Schaub and Fletcher (2015) take immigration to be proportional to the total population size of the observed population. Schaub and Fletcher (2015) also consider the case where immigration is taken to be a population-independent parameter, with one parameter for each year. The model of Abadi et al. (2010) simply adds the migration rate to the adult apparent survival probability, as we shall see in Equation 12, so that the model in effect remains a birth and death process, whereas in the alternative modelling of Schaub and Fletcher (2015), immigration is always present. A further possibility would be for the immigration rate to decrease as the population grows, for example through an appropriate logistic function to limit population size. Here we just use the model of Abadi et al. (2010), but alternative possibilities are easily explored.

Abadi et al. (2010) only analyse count data on female birds. They assume that breeding starts at age 1, and a balanced sex ratio at birth, so that in an obvious notation, their state equations, which only consider the female population, are given as:

$$N_{1,t+1} | \mathbf{N}_t \sim \text{Pois}(N_t r_t \phi_{f,1,t} / 2), \quad (10)$$

$$N_{a,t+1} | \mathbf{N}_t \sim \text{Bin}(N_t, \phi_{f,a,t}) + \text{Pois}(N_t \gamma), \quad (11)$$

where  $N_{1,t}$  and  $N_{a,t}$  denote the numbers of one-year old female birds and (adult) female birds aged  $\geq 2$  years respectively at time  $t$ ,  $\gamma$  is the immigration rate,  $\mathbf{N}_t = (N_{1,t}, N_{a,t})$  and  $N_t = N_{1,t} + N_{a,t}$ , for  $t = 1, 2, \dots, T$ . In terms of a Leslie matrix for a Kalman filter analysis, we can write the state equations as

$$\begin{pmatrix} N_{1,t+1} \\ N_{a,t+1} \end{pmatrix} = \begin{pmatrix} r_t \phi_{f,1,t}/2 & r_t \phi_{f,1,t}/2 \\ \phi_{f,a,t} + \gamma & \phi_{f,a,t} + \gamma \end{pmatrix} \begin{pmatrix} N_{1,t} \\ N_{a,t} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{a,t} \end{pmatrix}, \quad (12)$$

where the additive binomial and Poisson error terms are specified above. The state process of Abadi et al. (2010) is thus two-dimensional. However adding Equations 10 and 11 gives

$$N_{t+1}|N_t \sim \text{Bin}(N_t, \phi_{f,a,t}) + \text{Pois}(N_t(\gamma + r_t \phi_{f,1,t}/2)). \quad (13)$$

We note that we obtain the same expression as that of Equation 13 if the  $\gamma$  terms appear in the first row of the Leslie matrix of Equation 12, rather than the second row, corresponding to making a different assumption for the unknown age of immigrant birds, respectively aged 1 year and at least 2 years. Should estimates of, eg.,  $\{N_{1,t}\}$ , be required, then they can be deduced from Equation 10.

The observation equation adopted by Abadi et al. (2010) is given by

$$y_t|N_t \sim \text{Pois}(N_t), \quad t = 1, \dots, T, \quad (14)$$

so that there is no separate variance for the observation equation in this case. In combination, Equations 13 and 14 specify a one-dimensional state-space model which we fit using HMMs. The elements of the transition probability matrices  $\{\mathbf{T}_t\}$  are the binomial-Poisson convolution probabilities of Equation 13, and the probability matrices  $\{\mathbf{P}(y_t)\}$  are diagonal matrices providing the Poisson probabilities from Equation 14. The likelihood of Equation 8 is easily programmed, and the potential complication of dealing with matrices of very large dimension does not arise in the case of the Little owl data. For the initial distribution vector  $\boldsymbol{\delta}$  of Equation 8 we assume a Poisson distribution over states.

For the Kalman-filter analysis, we approximate  $N_{t+1}|N_t \sim N(N_t(\phi_{f,a,t} + \gamma + r_t \phi_{f,1,t}/2), \omega^2)$ ,

176 with  $\omega^2 = N_t\{\phi_{f,a,t}(1 - \phi_{f,a,t}) + (\gamma + r_t\phi_{f,1,t}/2)\}$ , and take  $y_t|N_t \approx N(N_t, a_t)$ , where  $a_t$  is the  
177 one-step-ahead prediction from the Kalman filter; see McCrea and Morgan (2014, p. 214)  
178 and Newman et al. (2014, p.64).

## 179 4.2 *Little owl results*

180 We present in Figure 1 illustrative results from hidden Markov modelling, the Bayesian  
181 analysis, taken from Abadi et al. (2010), and from using the Kalman filter, when the  
182 productivity data are included in the integrated analysis. We note the close correspondence  
183 of the results from the three methods, and the large confidence intervals which suggest that  
184 survival probabilities can be taken as constant. The productivity estimates are essentially  
185 the values of  $\{\hat{r}_t = j_t/V_t\}$  given in Section 3.2.

186 [Figure 1 about here.]

187 [Figure 2 about here.]

188 Figure 2 shows that when productivity is taken as constant,  $r$ , and the productivity data  
189 are omitted from the integrated analysis then the log-likelihood surface possess a ridge. The  
190 model is parameter redundant, and it is not possible to estimate the immigration rate; the  
191 same feature applies when productivity varies with time. This can be verified formally using  
192 a modification of the Maple code associated with Cole and McCrea (2016). A check for weak  
193 identifiability in Bayesian modelling is provided by Gimenez et al. (2009), which might also  
194 be used in this context. Thus in the absence of productivity data, the parameter estimates  
195 for the productivity obtained by Abadi et al. (2010) can be seen to be driven primarily by  
196 the  $U(0, 5)$  prior used for the productivity: from the values presented in Appendix S2-C of  
197 Abadi et al. (2010), the time-averaged estimates of productivity mean and standard error are  
198 respectively 2.21 and 1.53, compared with 2.5 and 1.44 for the  $U(0, 5)$  prior distribution used.  
199 Appendix S2-B of Abadi et al. (2010) provides parameter estimates for when the productivity

200 data are included in the IPM. It is a coincidence that the average productivity is then 2.34.  
 201 We note here also an example in Barry et al. (2003), in which a likelihood surface with a  
 202 flat ridge and flat priors can result in univariate marginal posterior distributions that are  
 203 unimodal.

204 When analysing all of the data, Abadi et al. (2010) investigate alternative models for the  
 205 immigration parameter, considering whether it varies over time or varies with an indicator of  
 206 the presence of voles, on which Little owls prey. Using classical inference, AIC can be used,  
 207 and in Table 1 we compare a range of models for immigration also considered by Abadi et al.  
 208 (2010). We see that the best model has constant immigration, and the model with regression  
 209 on the vole indicator is a competitor; Abadi et al. (2010) drew a similar conclusion based on  
 210 the DIC.

211 [Table 1 about here.]

### 212 4.3 Northern lapwings

The state equations for Northern lapwings are taken from Besbeas et al. (2002)

$$N_{1,t+1} | \mathbf{N}_t \sim \text{Pois}(N_{a,t} r_t S_{1,t} / 2),$$

$$N_{a,t+1} | \mathbf{N}_t \sim \text{Bin}(N_{1,t} + N_{a,t}, S_{a,t}), \quad (15)$$

which for the Kalman filter analysis have the matrix formulation,

$$\begin{pmatrix} N_{1,t+1} \\ N_{a,t+1} \end{pmatrix} = \begin{pmatrix} 0 & r_t S_{1,t} / 2 \\ S_{a,t} & S_{a,t} \end{pmatrix} \begin{pmatrix} N_{1,t} \\ N_{a,t} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{a,t} \end{pmatrix}. \quad (16)$$

213 Here again,  $N_{1,t}$  and  $N_{a,t}$  denote the numbers of one-year old female birds and (adult) female  
 214 birds aged  $\geq 2$  years respectively at time  $t$ ,  $S_{1,t}$  and  $S_{a,t}$  are respectively the annual survival  
 215 probabilities of birds in their first year of life and of birds aged 1 year and older at time  $t$ ,  
 216 and  $r_t$  denotes productivity at time  $t$ . We assume no sex effect on survival and in this case  
 217 that breeding starts at age 2. It is clear from Equation 16 that in this case the state space

is two-dimensional. For hidden Markov modelling we shall assume that  $N_{1,t} \leq N_{1,max}$  and  $N_{a,t} \leq N_{a,max}$ , for all  $t$ , for values  $N_{1,max}$  and  $N_{a,max}$  which need to be determined.

We are not able to observe  $\{N_{1,t}\}$ , as information is available only on the numbers breeding,  $\{N_{a,t}\}$ , and the observation equation adopted by Besbeas et al. (2002) is given by

$$y_t | \mathbf{N}_t \sim N(N_{a,t}, \sigma^2),$$

where  $\sigma^2$  is a free parameter to estimate.

Eliminating  $\{N_{1,t}\}$  from Equation 15, we can see that a hidden process for  $\{N_{a,t}\}$  alone is a second-order Markov chain, with state-space dimension  $N_{a,max}^2$ , as

$$\begin{aligned} Pr(N_{a,t+1} = k | N_{a,t}, N_{a,t-1}) &= \sum_{j=0}^{\infty} \binom{j + N_{a,t}}{k} S_{a,t}^k (1 - S_{a,t})^{j + N_{a,t} - k} \times \\ &e^{(-N_{a,t-1} r_t S_{1,t} / 2)} (N_{a,t-1} r_t S_{1,t} / 2)^j / j!. \end{aligned}$$

This explains in part why it is more efficient to use the first-order chain, given below, when, as here, we have  $N_{1,max} < N_{a,max}$ . We therefore use the first-order chain with state vector the set of values taken by the ordered pair,  $(N_{1,t}, N_{a,t})$ . Thus the potential values taken by the state vector are:

$$(0, 0), (0, 1), \dots, (0, N_{a,max}), (1, 0), \dots, (1, N_{a,max}) \dots (N_{1,max}, 0), \dots, (N_{1,max}, N_{a,max})$$

and the entries of  $\mathbf{\Gamma}_t$  are

$$\Pr(N_{1,t+1} = w, N_{a,t+1} = x | N_{1,t} = u, N_{a,t} = v) = \exp^{-\lambda_v} \lambda_v^w / w! \times \binom{u + v}{x} S_{a,t}^x (1 - S_{a,t})^{u + v - x}, \quad (17)$$

where  $\lambda_v = v r_t S_{1,t} / 2$ ,  $w = 0, 1, 2, \dots$ ,  $x = 0, 1, \dots, u + v$ .

The matrix  $\mathbf{\Gamma}_t$  is a partitioned matrix with the block structure given below:

$$\mathbf{\Gamma}_t = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \cdots & \mathbf{A}_{1,N_{1,max},1} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \cdots & \mathbf{A}_{2,N_{1,max},1} \\ \vdots & & \ddots & \vdots \\ \mathbf{A}_{N_{1,max},1} & \mathbf{A}_{N_{1,max},2} & \cdots & \mathbf{A}_{N_{1,max},N_{1,max}} \end{bmatrix},$$

where each submatrix has dimension  $N_{a,max} \times N_{a,max}$ . We now describe the submatrices.

Conditional upon  $N_{1,t} = u$ , and  $N_{a,t} = v$ , the appropriate submatrices of  $\mathbf{\Gamma}_t$  are those that comprise the  $u^{th}$  row, and the probabilities of Equation 17 form the entries of the  $v^{th}$  rows of the sub matrices, with  $w = i$  for the  $i^{th}$  column sub matrices. Apart from the first term in Equation 17, these rows are identical, and the only difference arises from the Poisson probability multiplier in Equation 17. Computationally, it is convenient to express  $\mathbf{\Gamma}_t$  in terms of ordinary and nested Kronecker product operations of binomial and Poisson probabilities as follows, where we suppress the time dependence for convenience.

Define

$$\mathbf{B}_u = \left\{ \binom{u+v}{x} S_a^x (1 - S_a)^{u+v-x} \right\}_{N_{a,max} \times N_{a,max}},$$

for

$$v = 0, \dots, N_{a,max}, \quad x = 0 \dots, N_{a,max} \quad u = 0, \dots, N_{1,max},$$

and set

$$\mathbf{B} = \mathbf{1} \otimes \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \vdots \\ \mathbf{B}_{N_{1,max}} \end{bmatrix},$$

where  $\mathbf{1}$  is now a  $1 \times N_{1,max}$  row vector of 1s, and  $\otimes$  is the Kronecker product operator.

Define the column vector  $\mathbf{R}_w$  by

$$\mathbf{R}_w = \left\{ \frac{e^{-\lambda_v} \lambda_v^w}{w!} \right\}_{N_{a,max} \times 1},$$

for

$$v = 0, 1, \dots, N_{a,max}, \quad w = 1, \dots, N_{1,max},$$

and let

$$\mathbf{C}_w = \mathbf{1}^\top \otimes (\mathbf{1} \otimes \mathbf{R}_w), \quad \text{and}$$

$$\mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_{N_{1,max}}].$$

Then in terms of the Hadamard product we can write,

$$\mathbf{\Gamma} = \mathbf{C} \circ \mathbf{B}. \quad (18)$$

The  $\mathbf{P}(y_t)$  are diagonal,  $(N_{1,max}N_{a,max} \times N_{1,max}N_{a,max})$  matrices with appropriate entries for the normal probability density function of the observations,  $\{y_t|N_{a,t}\}$ , replicated for each of the sub matrices of  $\mathbf{\Gamma}$ . Thus we can write  $\mathbf{P} = \mathbf{I}_{N_{1,max}} \otimes \mathbf{Q}$ , where  $\mathbf{Q}$  is a diagonal  $N_{a,max} \times N_{a,max}$  matrix containing the probability density terms for  $\{y_t|N_{a,t}\}$ .

#### 4.4 The use of binning

In contrast to the Little owl example, as a consequence of the binomial index of Equation 15, the state vector is a one-dimensional vector of size  $N_{1,max}N_{a,max}$ . In the lapwing application we take  $N_{1,max} = 800$  and  $N_{a,max} = 2200$ , following experimentation; in such a situation we use binning to group elements of the state vector to reduce its size, and it is inefficient to use the same bin widths for the different age components of the state vector. Note that for computing  $\mathbf{\Gamma}_t$ , for rows we take the mid points of bins, whereas for columns we appropriately use the cumulative distribution function for the appropriate discrete distributions. In the lapwing example, the results from Besbeas et al. (2002), obtained from using the Kalman filter, demonstrate that the estimates of  $\{N_{1,t}\}$  are generally far smaller than those for  $\{N_{a,t}\}$ . This suggests using more bins for the adult age class than for the younger one. For the heron data analysis of Besbeas et al. (2002) there were three age classes: again the oldest age class has the largest estimated numbers, and the same consideration applies. A preliminary analysis for these models is straightforward, for example using the approximate Kalman filter approach or using a time-homogeneous HMM. This can suggest the use of differential bin sizes according to component of the state vector. A further approach would be to have several bin widths within each age class, with widths increasing with distance from the values estimated from the Kalman filter analysis.

253 4.5 *Lapwing results*

254 The HMM modelling of the lapwing data is more complex than that of the Little owl data,  
255 as the model is two-dimensional, and population sizes are appreciably larger. The approach  
256 of using a second-order Markov chain (Zucchini et al., 2016, p. 148) was found to give the  
257 same results as using a first-order chain, but to be substantially less efficient, as anticipated  
258 above. The survival parameters are taken as  $\text{logit}(S_{1,t}) = \alpha_0 + \alpha_1 c_t$  and  $\text{logit}(S_{a,t}) = \beta_0 +$   
259  $\beta_1 c_t$ . Similarly, the reporting probability of dead birds is logistically regressed on time, with  
260 parameters  $\gamma_0, \gamma_1$ , and productivity is logarithmically regressed on time, with parameters  
261  $\delta_0, \delta_1$ . We present the results of several analyses using binning, as well as Kalman filter  
262 and Bayesian results, in Table 2. Two Kalman filter approaches are used, one requiring  
263 initialisation and the other including maximum-likelihood estimation of the initial state  
264 (Besbeas and Morgan, 2010), which is directly comparable to the HMM approach. Four  
265 values of the bin width are used, viz.,  $w = 10, 20, 40, 50$ . We can see that the smallest values  
266 of  $w$  generally result in virtually identical estimates to using the Kalman filter. However  
267 there is little difference between the different analyses, which agree also with the Bayesian  
268 results of Brooks et al. (2004).

269 If we take two different bin widths for the two age classes, then we can denote the bin  
270 widths as  $w_1$  and  $w_a$ , for the aged 1 and older age classes respectively. Then for example,  
271 if we take  $w_1 = 20$  and  $w_a = 40$ , the dimension of the state vector is 2296, compared with  
272 values of 765, 4551 and 17901 when  $w_1 = w_a = 50$ ,  $w_1 = w_a = 20$  and  $w_1 = w_a = 10$ ,  
273 respectively.

274 [Table 2 about here.]

275 We can see from Table 3 how changing the bin widths can affect the computation time for  
276 a likelihood evaluation. In this example, with larger values for the numbers of adult birds it

is best to take  $w_a > w_1$ . As shown in Table 2, below a certain value, changing bin widths has little effect on parameter estimates.

We can test for additive vs multiplicative errors in the observation equation quite easily using the HMM format. In this application, taking  $w_1 = w_a = 20$ , assuming normal errors results in a log-likelihood maximum value of -7379.4, compared with -7380.3 for the lognormal case. Changing  $w_1$  and  $w_a$  makes no appreciable change to this comparison, and we see here that there is little difference between these models for this application.

[Table 3 about here.]

## 5. Discussion and future research

### 5.1 State-space dimension

Much IPM will be one-dimensional; see for example the models of Baillie et al. (2009) and Robinson et al. (2014), of wide-ranging importance for typical long-term data on short-lived species. The same is true of models for seasonal insects, see Freeman (2009), and models commonly used in fisheries, for example using a Gompertz model; see Knappe et al. (2011). Analysis of population time-series alone in such cases using HMM is the topic of Besbeas and Morgan (2018). As we have seen in modelling Little owl data, the dimensionality of a state space may be reduced. Besbeas et al. (2002) adopt the following transition equation for Grey herons, *Ardea cinerea*

$$\begin{pmatrix} N_{1,t+1} \\ N_{2,t+1} \\ N_{a,t+1} \end{pmatrix} = \begin{pmatrix} 0 & rS_{1,t}/2 & rS_{1,t}/2 \\ S_{2,t} & 0 & 0 \\ 0 & S_{a,t} & S_{a,t} \end{pmatrix} \begin{pmatrix} N_{1,t} \\ N_{2,t} \\ N_{a,t} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{2,t} \\ \eta_{a,t} \end{pmatrix}, \quad (19)$$

where  $N_{2,t}$  is the number of birds of age 2 at time  $t$  and  $N_{a,t}$  denotes the number of birds aged  $\geq 3$  at time  $t$ . We can see that here too the dimension of the state space can be reduced

in size by one. Whilst it will depend on the age of breeding assumed, this simplifying feature will commonly be the case. See for example Finke et al. (2019) for a further illustration.

The extension of Equation 18 to the case of more than 2 age classes is in principle straightforward, and dependent on the specifics of the Leslie matrix used in the model.

## 5.2 Numerical choices and potential

The HMM approach opens the way to using standard likelihood tools, to check for parameter redundancy (Cole et al., 2010), goodness-of-fit (Besbeas and Morgan, 2014), over dispersion, to perform model selection (Besbeas et al., 2015), to include non-linearity, for example to describe density-dependence, and to compare the performance of alternative distributions, as for the observation equation case in the last section; cf Knappe et al. (2011). The only costs are those of deciding on a suitable size(s) of bin width when binning is needed, and on the maximum length(s) for the state vector, which can be obtained experimentally. Thus in comparison with the Kalman-filter approach, we are in effect replacing statistical approximations with numerical ones. For state spaces of dimension  $> 1$ , binning will probably be necessary. For dimensions  $> 2$ , then ways of speeding up the HMM approach may be necessary, for example by combining the bin-width selection procedures that we suggest, and exploiting the sparse structure of the  $\Gamma_t$  matrices. This is a promising research area.

## 6. Supplementary materials

A zip file containing MATLAB<sup>®</sup> program files, a README file and an illustrative example are available with this paper at the *Biometrics* website on Wiley Online Library.

**Acknowledgements** We thank the editor, associate editor and two referees for their positive and helpful comments, and Diana Cole for confirming parameter redundancy for the Little owl analysis. We acknowledge those responsible for the CBC data. The CBC was supported by the BTO and the Joint Nature Conservation Committee. PTB was partly

321 supported by an Original Research Grant, AUEB. BJTM was supported by a Leverhulme  
322 Emeritus Fellowship.

## 323 References

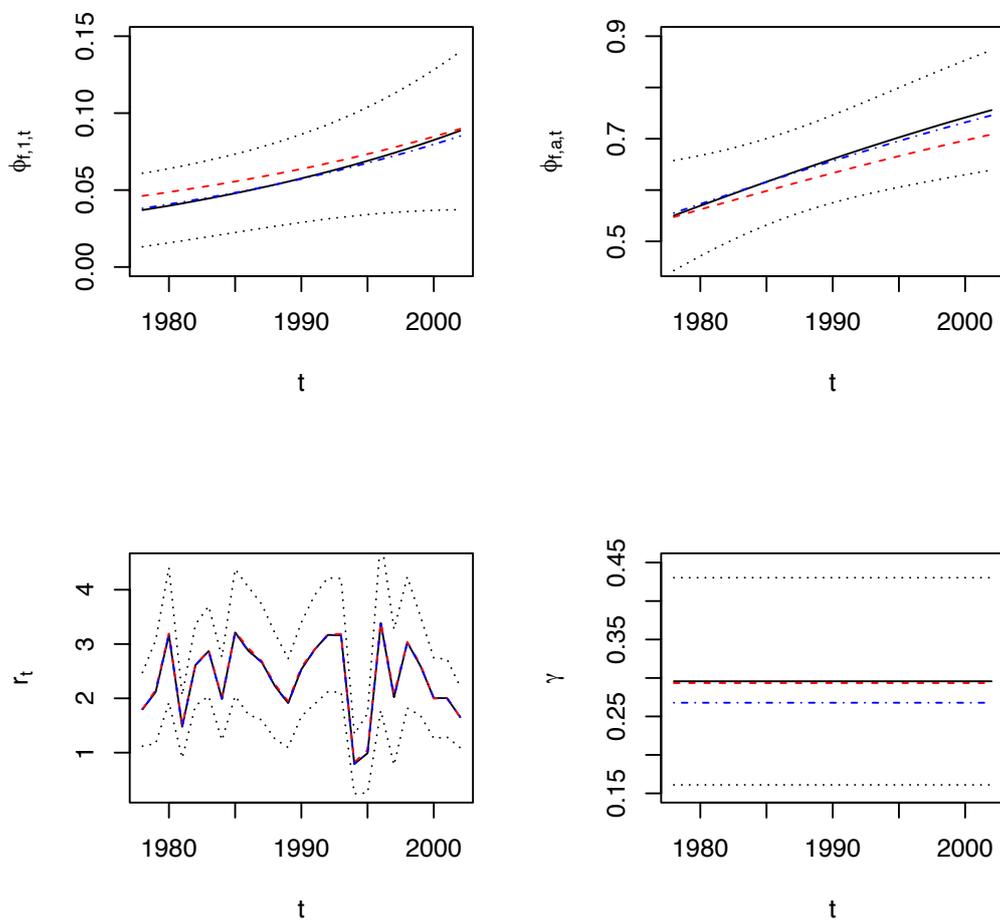
- 324 Abadi, F., Gimenez, O., Ullrich, B., Arlettaz, R., and Schaub, M. (2010). Estimation of  
325 immigration rate using integrated population models. *Journal of Applied Ecology* **47**,  
326 393–400.
- 327 Baillie, S. R., Brooks, S. P., King, R., and Thomas, L. (2009). Using a state-space model  
328 of the British song thrush *Turdus philomelos* population to diagnose the causes of a  
329 population decline. In Thomson, D. L., Cooch, E. G., and Conroy, M. J., editors,  
330 *Modelling Demographic Processes in Marked Populations*, pages 541–561. Springer.
- 331 Barry, S. C., Brooks, S. P., Catchpole, E. A., and Morgan, B. J. T. (2003). The analysis of  
332 ring-recovery data using random effects. *Biometrics* **59**, 54–65.
- 333 Besbeas, P., Freeman, S. N., Morgan, B. J. T., and Catchpole, E. A. (2002). Integrating mark-  
334 recapture-recovery and census data to estimate animal abundance and demographic  
335 parameters. *Biometrics* **58**, 540–547.
- 336 Besbeas, P., McCrea, R. S., and Morgan, B. J. T. (2015). Integrated population model  
337 selection in ecology. Technical report, University of Kent, Canterbury CT2 7FS, England.
- 338 Besbeas, P. and Morgan, B. J. T. (2010). Kalman filter initialisation for integrated population  
339 modelling. *Applied Statistics* **61**, 151–162.
- 340 Besbeas, P. and Morgan, B. J. T. (2012). A threshold model for heron productivity. *Journal*  
341 *of Agricultural, Biological, and Environmental Statistics* **17**, 128–141.
- 342 Besbeas, P. and Morgan, B. J. T. (2014). Goodness of fit of integrated population models  
343 using calibrated simulation. *Methods in Ecology and Evolution* **5**, 1373–1382.
- 344 Besbeas, P. and Morgan, B. J. T. (2017). Variance estimation for integrated population  
345 models. *AStA Advances in Statistical Analysis* **101**, 1–22.

- 346 Besbeas, P. and Morgan, B. J. T. (2018). A general framework for modelling population  
347 abundance data: paper under revision. Technical report, University of Kent, Canterbury  
348 CT2 7FS, England.
- 349 Brooks, S. P., King, R., and Morgan, B. J. T. (2004). A Bayesian approach to combining  
350 animal abundance and demographic data. *Animal Biodiversity and Conservation* **27**,  
351 515–529.
- 352 Cole, D. J. and McCrea, R. S. (2016). Parameter redundancy in discrete state-space and  
353 integrated models. *Biometrical Journal* **5**, 1071–1090.
- 354 Cole, D. J., Morgan, B. J. T., and Titterton, D. M. (2010). The parametric structure of  
355 models. *Mathematical Biosciences* **228**, 16–30.
- 356 Cowen, L., Besbeas, P. T., Morgan, B. J. T., and Schwarz, C. (2017). Hidden Markov models  
357 for extended batch data. *Biometrics* page DOI: 10.1111/biom.12701.
- 358 deValpine, P. (2012). Frequentist analysis of hierarchical models for population dynamics  
359 and demographic data. *Journal of Ornithology* **152**, Supplement 2, S393–S408.
- 360 Finke, A., King, R., Beskos, A., and Dellaportas, P. (2019). Efficient sequential Monte Carlo  
361 algorithms for integrated population models. *Journal of Agricultural, Biological and  
362 Environmental Statistics* <https://doi.org/10.1007/s13253-018-00349-9>.
- 363 Freeman, S. (2009). Towards a method for the estimation and use of averaged multi-species  
364 trends, as indicators of patterns of change in butterfly populations. Technical report,  
365 Centre for Ecology and Hydrology, Wallingford, OX10 8BB, England.
- 366 Gimenez, O., Morgan, B. J. T., and Brooks, S. P. (2009). Weak identifiability in models for  
367 mark-recapture-recovery data. In Thomson, D. L., Cooch, E. G., and Conroy, M. J.,  
368 editors, *Modelling Demographic Processes in Marked Populations*. Springer.
- 369 Kéry, M. and Schaub, M. (2012). *Bayesian Population Analysis using WinBUGS: A  
370 Hierarchical Perspective*. Academic Press.

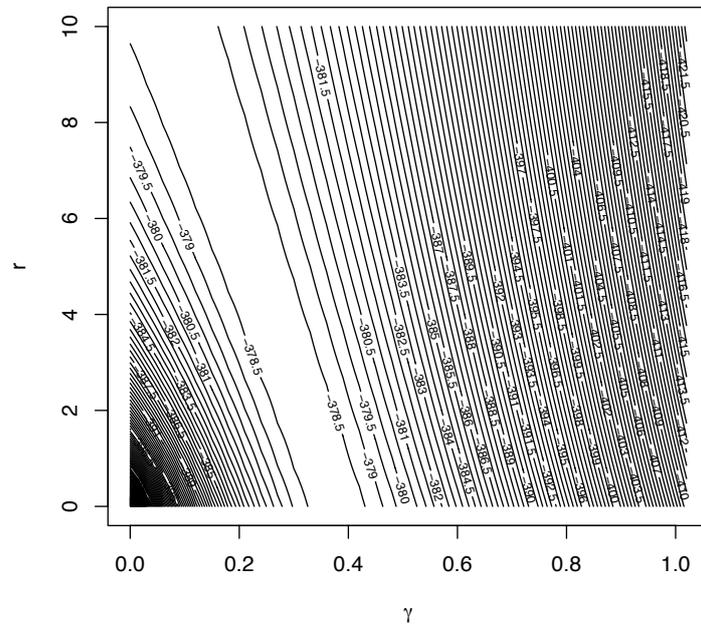
- 371 King, R. (2012). A review of Bayesian state-space modelling of capture-recapture data.  
372 *Interface Focus* **2**, 190–204.
- 373 King, R. (2014). Statistical ecology. *Annual Review of Statistics and Its Application* **1**,  
374 401–426.
- 375 Knape, J., Jonzén, and Sköld, M. (2011). On observation distributions for state space models  
376 of population survey data. *Journal of Animal Ecology* **80**, 1269–1277.
- 377 Lahoz-Monfort, J. J., Harris, M. P., Wanless, S., Freeman, S. N., and Morgan, B. J. T. (2017).  
378 Bringing it all together: multi-species integrated population modelling of a breeding  
379 community. *JABES* **22**, 140–160.
- 380 Maunder, M. N. and Punt, A. E. (2013). A review of integrated analysis in fisheries stock  
381 assessment. *Fisheries Research* **142**, 61–74.
- 382 McCrea, R. S. and Morgan, B. J. T. (2014). *Analysis of Capture-recapture Data*. CRC Press,  
383 Chapman & Hall, Boca Raton.
- 384 Newman, K. B., Buckland, S. T., Morgan, B. J. T., King, R., Borchers, D. L., Cole, D. J.,  
385 Besbeas, P., Gimenez, O., and Thomas, L. (2014). *Modelling Population Dynamics*.  
386 Springer, New York.
- 387 Robinson, R. A., Morrison, C. A., and Baillie, S. R. (2014). Integrating demographic data:  
388 towards a framework for monitoring wildlife populations at large spatial scales. *Methods*  
389 *in Ecology and Evolution* **5**, 1361–1372.
- 390 Schaub, M. and Abadi, F. (2011). Integrated population models: a novel analysis framework  
391 for deeper insights into population dynamics. *Journal of Ornithology* **152**, 227–237.
- 392 Schaub, M. and Fletcher, D. (2015). Estimating immigration using a Bayesian integrated  
393 population model: choices of parametrization and priors. *Environmental and Ecological*  
394 *Statistics* **22**, 535–549.
- 395 Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time*

Accepted Article

**Figure 1.** Results from IPM: comparison of Bayesian, Kalman filter and exact HMM analyses of the Little owl data. All three data sets are included in the integrated population model. Black denotes results from HMM, dashed lines are results from the Bayesian analysis and dashed and dotted lines are the results from using the Kalman filter (KF). Also shown, with dotted lines, are 95 % confidence bands from the HMM. The figure appears in colour online.



**Figure 2.** Two-parameter profile log likelihood for the Little owl data, showing the likelihood surface ridge, when the data on productivity are not included in the HMM analysis. For this analysis productivity is taken as constant,  $r$ .



**Table 1**

*Fitting Little owl data using the HMM approach: statistics from fitting 5 different models to investigate regressions of the immigration rate,  $\gamma$ , on time and on the vole indicator variable;  $\ell$  denotes the maximised log-likelihood value;  $\Delta AIC$  denotes the change in the Akaike information criterion (AIC) compared with the model with the smallest AIC value.*

model for $\gamma$	$-\ell$	$\Delta AIC$
constant	410.8	0.0
linear regression on vole indicator	410.5	1.4
linear regression on year	410.6	3.6
quadratic regression on year	410.0	4.4
full time dependence	405.8	58.0

**Table 2**

Maximum-likelihood parameter estimates corresponding to different ways of fitting a model to the lapwing data. The “Diffuse” results are taken from Besbeas et al. (2002), when the Kalman filter used a vague prior for the initial population sizes. The “MLE KF” results follow from using the Kalman filter with maximum-likelihood estimation of the initial population sizes,  $N_1, N_a$ . The hidden Markov modelling results are for the the bin widths shown; see text for details. Estimated standard errors are indicated by SE. The “Bayes” results are taken from Brooks et al. (2004), suitably adjusted for the different scaling of the weather covariates used, and in that case the values for  $\hat{N}_1$  and  $\hat{N}_a$  are estimated from Figure 4 of that paper; when available, estimated standard deviations are indicated by SD.

Model	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}_0$	$\hat{\gamma}_1$	$\hat{\delta}_0$	$\hat{\delta}_1$	$\hat{\sigma}$	$\log(\hat{N}_1)$	$\log(\hat{N}_a)$
Diffuse KF	0.523	-0.023	1.521	-0.028	-4.562	-0.584	-1.151	-0.432	159.469		
SE	0.067	0.007	0.069	0.005	0.035	0.064	0.088	0.074	21.870		
MLE KF	0.523	-0.023	1.521	-0.028	-4.563	-0.584	-1.178	-0.425	155.867	5.966	7.015
SE	0.068	0.007	0.070	0.005	0.035	0.064	0.091	0.076	21.198	0.546	0.135
$w_1 = w_a = 50$	0.520	-0.023	1.504	-0.028	-4.566	-0.582	-1.156	-0.406	152.170	5.941	7.803
SE	0.068	0.008	0.067	0.004	0.035	0.064	0.094	0.078	21.456	0.250	0.082
$w_1 = w_a = 40$	0.520	-0.023	1.509	-0.028	-4.565	-0.583	-1.165	-0.417	153.961	5.956	7.021
SE	0.067	0.007	0.069	0.005	0.035	0.064	0.091	0.076	21.332	0.577	0.144
$w_1 = w_a = 20$	0.523	-0.023	1.520	-0.028	-4.563	-0.584	-1.181	-0.427	155.711	5.966	7.014
SE	0.068	0.007	0.069	0.005	0.035	0.064	0.091	0.077	21.332	0.557	0.140
$w_1 = w_a = 10$	0.523	-0.023	1.520	-0.028	-4.563	-0.584	-1.182	-0.427	156.179	5.965	7.014
SE	0.067	0.007	0.069	0.005	0.035	0.064	0.091	0.076	21.412	0.577	0.145
Bayes	0.543	-0.024	1.550	-0.029	-4.522	-0.578	-1.154	-0.459	169.112	6.016	7.003
SD	0.067	0.007	0.070	0.005	0.035	0.069	0.089	0.079	23.001		

**Table 3**

*A comparison of timings, in seconds, for a likelihood evaluation for the lapwing analysis, as bin widths  $w_1$  and  $w_a$  vary.*

	$w_a$				
	50	40	20	10	
$w_1$	50	1.17	1.72	6.23	21.75
	40	1.41	1.92	8.24	27.18
	20	3.48	5.31	17.30	57.09
	10	7.98	11.45	38.09	135.39