Received: 27 February 2018 | Accepted: 6 December 2018
DOI: 10.1111/biom.13045

ORIGINAL ARTICLE

WILEY Biometrics

Exact inference for integrated population modelling

P. Besbeas[1,2] | **B.J.T. Morgan**[2]

[1]Department of Statistics, Athens University of Business and Economics, 10434 Athens, Greece

[2]National Centre for Statistical Ecology, School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, Kent CT2 7FS, England

Correspondence
P. Besbeas, Department of Statistics, Athens University of Business and Economics, 10434 Athens, Greece
Email: P.T.Besbeas@kent.ac.uk

Abstract

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.

KEYWORDS

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 article 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 Knape 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 article. 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 article 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 to 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 and 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 version of the covariate, are embedded in the WinBUGS code provided by Brooks et al. (2004).

3 | COMPONENT AND INTEGRATED MODELLING

Throughout we use boldface to indicate generic parameters which may involve several coefficients due to variation over time and/or by age.

3.1 | Survival

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

We denote the likelihood for capture-recapture data as $L_C(\boldsymbol{\phi}, \boldsymbol{p}; \boldsymbol{m})$, in which we use $\boldsymbol{m}$ to denote the matrix of numbers of recaptures, commonly called the m-array (McCrea and Morgan, 2014, p. 69). The corresponding notation adopted for the likelihood for recovery data is $L_R(\boldsymbol{S}, \boldsymbol{\lambda}; \boldsymbol{d})$, where $\boldsymbol{d}$ is the matrix of numbers of recoveries. For both matrices, each row contains the recorded numbers for the year of release of marked individuals corresponding to that row, and the columns indicate the year of recovery or recapture, respectively. The vector $\boldsymbol{R}$ provides the annual totals of marked birds released. For illustration we provide 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 ϕ_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:

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

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 $1 \le i < T$, and $\chi_T = 1$. The product-multinomial likelihood is then given by

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

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 ϕ_{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-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

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

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

3.2 | Productivity

Abadi et al. (2010) adopt yearly estimation of time-dependent model fecundities, assumed to have Poisson distributions, $\{J_t\} \sim \mathrm{Pois}(V_t r_t)$, 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

$$\mathrm{L}_P(\boldsymbol{r}; \boldsymbol{j}, \boldsymbol{V}) \propto \prod_{t=1}^{T} e^{-V_t r_t} (V_t r_t)^{j_t},$$

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 $\boldsymbol{N}_t = (N_{1,t}, \ldots, N_{K,t})'$, for $t = 1, \ldots, T$, where $N_{j,t}$ is the number of individuals in state $1 \le j \le K$ at time t, and the annual population counts, $\{\mathbf{y}_t\}$, form an M-variate time series, for $M \ge 1$. The general formulation for state-space models links the state and observation processes as follows (Newman et al., 2014, p. 43)

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

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

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

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

We write the likelihood for the time-series data when survival estimation is based upon capture-recapture data as $\mathrm{L}_T(\boldsymbol{\phi}, \boldsymbol{r}, \boldsymbol{\sigma}, \boldsymbol{N}_1; \boldsymbol{y}_t)$, if variance parameters $\boldsymbol{\sigma}$ for the observation equation and $\boldsymbol{N}_1$ 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,

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

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

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

In addition random variables appearing in the variance terms in the state equation (6) are approximated by their expectations. The likelihood is easily formed, the method is fast, performs well and is robust with respect to departures from the assumptions and approximations made, even for small population sizes; see Brooks et al. (2004).

3.3.2 | Hidden Markov models

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

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

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

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

3.3.3 | Bayesian inference

The Bayesian approach uses MCMC and also does not need 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

$$\mathrm{L}_I(\boldsymbol{r}, \boldsymbol{\phi}, \boldsymbol{\sigma}, \boldsymbol{p}, \boldsymbol{N}_1; \boldsymbol{j}, \boldsymbol{V}, \boldsymbol{y}_t, \boldsymbol{m}) = \mathrm{L}_P(\boldsymbol{r}; \boldsymbol{j}, \boldsymbol{V})\mathrm{L}_T(\boldsymbol{\phi}, \boldsymbol{r}, \boldsymbol{\sigma}, \boldsymbol{N}_1; \boldsymbol{y}_t)\mathrm{L}_C(\boldsymbol{\phi}, \boldsymbol{p}; \boldsymbol{m}), \tag{9}$$

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, $\mathrm{L_I}$, 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 $\boldsymbol{N}_1$ and/or $\boldsymbol{\sigma}$ 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:

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

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

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

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:

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

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

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

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

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

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

We note that we obtain the same expression as that of equation (13) if the γ 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

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

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 $\{\boldsymbol{\Gamma}_t\}$ are the binomial-Poisson convolution probabilities of equation (13), and the probability matrices $\{\boldsymbol{P}(y_t)\}$ 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 $N_{t+1}|N_t \sim N(N_t(\phi_{f,a,t} + \gamma + r_t\phi_{f,1,t}/2), \omega^2)$, with $\omega^2 = N_t\{\phi_{f,a,t}(1 - \phi_{f,a,t}) + (\gamma + r_t\phi_{f,1,t}/2)\}$, and take $y_t|N_t \approx \text{N}(N_t, a_t)$, where a_t is the one-step-ahead prediction from the Kalman filter; see McCrea and Morgan (2014, p. 214) and Newman et al. (2014, p. 64).

4.2 | Little owl results

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

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

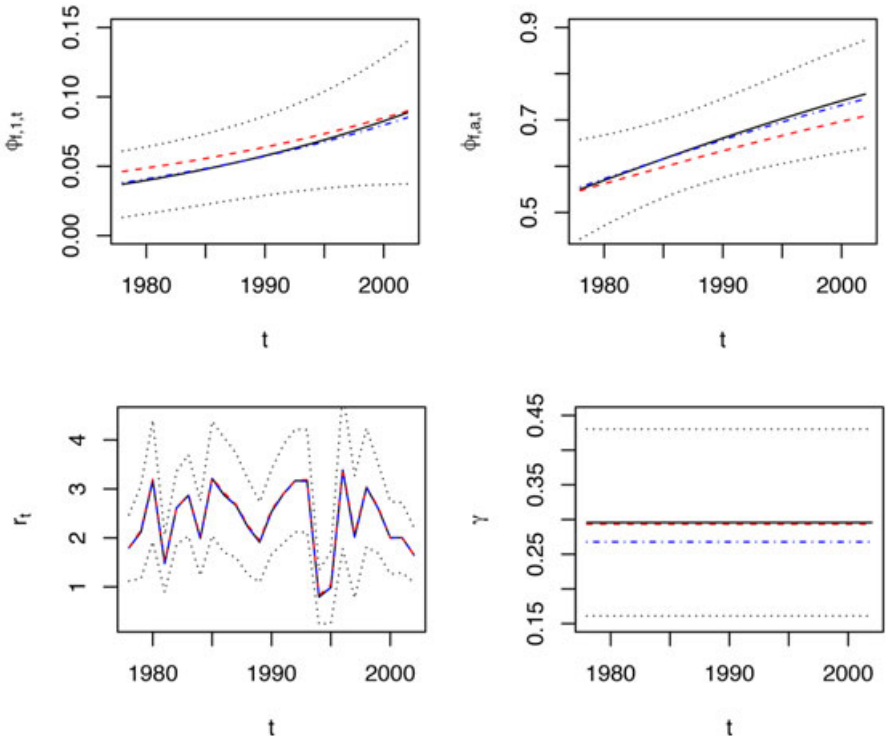

FIGURE 1 Results from IPM: comparison of Bayesian, Kalman filter and exact HMM analyses of the Little owl data. All three data sets are included in the integrated population model. Black denotes results from HMM, dashed lines are results from the Bayesian analysis and dashed and dotted lines are the 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.

al. (2010) provides parameter estimates for when the productivity data are included in the IPM. It is a coincidence that the average productivity is then 2.34. We note here also an example in Barry et al. (2003), in which a likelihood surface with a flat ridge and flat priors can result in univariate marginal posterior distributions that are unimodal.

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

4.3 | Northern lapwings

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

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

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

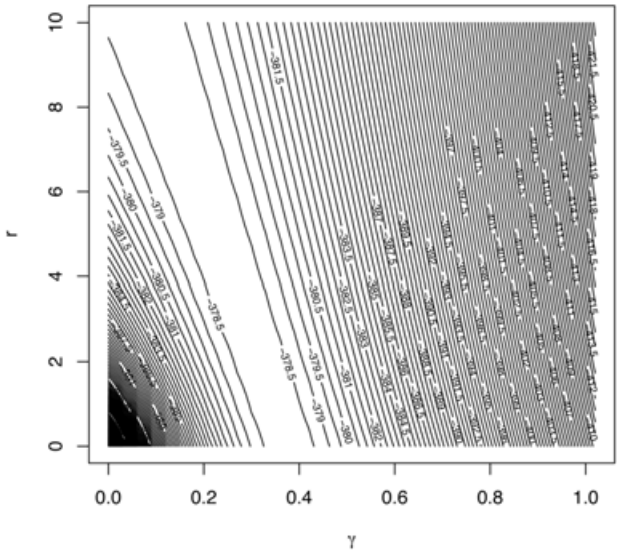

FIGURE 2 Two-parameter profile log likelihood for the Little owl data, showing the likelihood surface ridge, when the data on productivity are not included in the HMM analysis. For this analysis productivity is taken as constant, r.

which for the Kalman filter analysis have the matrix formulation,

$$\begin{pmatrix} N_{1,t+1} \\ N_{a,t+1} \end{pmatrix} = \begin{pmatrix} 0 & r_tS_{1,t}/2 \\ S_{a,t} & S_{a,t} \end{pmatrix} \begin{pmatrix} N_{1,t} \\ N_{a,t} \end{pmatrix} + \begin{pmatrix} \eta_{1,t} \\ \eta_{a,t} \end{pmatrix}. \tag{16}$$

TABLE 1 Fitting Little owl data using the HMM approach: statistics from fitting 5 different models to investigate regressions of the immigration rate, γ, on time and on the vole indicator variable; ℓ denotes the maximised log-likelihood value; ΔAIC denotes the change in the Akaike information criterion (AIC) compared with the model with the smallest AIC value.

Model for γ	$-\ell$	ΔAIC
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

Here again, $N_{1,t}$ 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, $S_{1,t}$ and $S_{a,t}$ are respectively the annual survival probabilities of birds in their first year of life and of birds aged 1 year and older at time t, and r_t denotes productivity at time t. We assume no sex effect on survival and in this case 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

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

where σ^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 $N^2_{a,max}$, 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:

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

and the entries of $\boldsymbol{\Gamma}_t$ 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 $\lambda_v = v r_t S_{1,t}/2, \quad w = 0,1,2,\ldots, \quad x = 0,1,\ldots,u+v$.

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

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

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 $\boldsymbol{\Gamma}_t$ 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 $\boldsymbol{\Gamma}_t$ 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

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

for

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

and set

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

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

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

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

for

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

and let

$$\boldsymbol{C_w} = \mathbf{1}^{\intercal} \otimes (\mathbf{1} \otimes \boldsymbol{R_w}), \quad \text{and}$$

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

Then in terms of the Hadamard product we can write,

$$\boldsymbol{\Gamma} = \boldsymbol{C} \circ \boldsymbol{B}. \tag{18}$$

The $\boldsymbol{P}(y_t)$ 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 $\boldsymbol{\Gamma}$. Thus we can write $\boldsymbol{P} = \boldsymbol{I}_{N_{1,max}} \otimes \boldsymbol{Q}$, where $\boldsymbol{Q}$ 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 $\boldsymbol{\Gamma}_t$, 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.

4.5 | Lapwing results

The HMM modelling of the lapwing data is more complex than that of the Little owl data, as the model is two-dimensional, and population sizes are appreciably larger. The approach of using a second-order Markov chain (Zucchini et al., 2016, p. 148) was found to give the same results as using a first-order chain, but to be substantially less efficient, as anticipated above. The survival parameters are taken as $\text{logit}(S_{1,t}) = \alpha_0 + \alpha_1 c_t$ and $\text{logit}(S_{a,t}) = \beta_0 + \beta_1 c_t$. Similarly, the reporting probability of dead birds is logistically regressed on time, with parameters γ_0, γ_1, and productivity is logarithmically regressed on time, with parameters δ_0, δ_1. We present the results of several analyses using binning, as well as Kalman filter and Bayesian results, in Table 2. Two Kalman filter approaches are used, one requiring initialisation and the other including maximum-likelihood estimation of the initial state (Besbeas and Morgan, 2010), which is directly comparable to the HMM approach. Four values of the bin width are used, viz., $w = 10, 20, 40, 50$. We can see that the smallest values of w generally result in virtually identical estimates to using the Kalman filter. However there is little difference between the different analyses, which agree also with the Bayesian results of Brooks et al. (2004).

If we take two different bin widths for the two age classes, then we can denote the bin widths as w_1 and w_a, for the aged 1 and older age classes respectively. Then for example, if we take $w_1 = 20$ and $w_a = 40$, the dimension of the state vector is 2296, compared with values of 765, 4551 and 17,901 when $w_1 = w_a = 50$, $w_1 = w_a = 20$ and $w_1 = w_a = 10$, respectively.

We can see from Table 3 how changing the bin widths can affect the computation time for 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.

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 Knape et al. (2011). Analysis of population time-series alone in such cases using HMM is the topic of Besbeas and Morgan (2018). As we

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 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 article; 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

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*

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

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 ≥ 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 Knape 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 $\mathbf{\Gamma}_t$ matrices. This is a promising research area.

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

ORCID

B.J.T. Morgan http://orcid.org/0000-0002-5465-8006

REFERENCES

Abadi, F., Gimenez, O., Ullrich, B., Arlettaz, R., and Schaub, M. (2010). Estimation of immigration rate using integrated population models. *J Appl Ecol* 47, 393–400.

Baillie, S. R., Brooks, S. P., King, R., and Thomas, L. (2009). Using a state-space model of the British song thrush *Turdus philomelos* population to diagnose the causes of a population decline. In Thomson, D. L., Cooch, E. G., and Conroy, M. J., Eds, *Modelling Demographic Processes in Marked Populations*, pages 541–561. Springer.

Barry, S. C., Brooks, S. P., Catchpole, E. A., and Morgan, B. J. T. (2003). The analysis of ring-recovery data using random effects. *Biometrics* 59, 54–65.

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, 540–547.

Besbeas, P., McCrea, R. S., and Morgan, B. J. T. (2015). Integrated population model selection in ecology. Technical report, University of Kent, Canterbury CT2 7FS, England.

Besbeas, P. and Morgan, B. J. T. (2010). Kalman filter initialisation for integrated population modelling. *Appl Stat* 61, 151–162.

Besbeas, P. and Morgan, B. J. T. (2012). A threshold model for heron productivity. *J Agric Biol Environ Stat* 17, 128–141.

Besbeas, P. and Morgan, B. J. T. (2014). Goodness of fit of integrated population models using calibrated simulation. *Methods Ecol Evol* 5, 1373–1382.

Besbeas, P. and Morgan, B. J. T. (2017). Variance estimation for integrated population models. *Adv Stat Anal* 101, 1–22.

Besbeas, P. and Morgan, B. J. T. (2018). A general framework for modelling population abundance data: paper under revision. Technical report, University of Kent, Canterbury CT2 7FS, England.

Brooks, S. P., King, R., and Morgan, B. J. T. (2004). A Bayesian approach to combining animal abundance and demographic data. *Anim Biodivers Conserv* 27, 515–529.

Cole, D. J. and McCrea, R. S. (2016). Parameter redundancy in discrete state-space and integrated models. *Biom J* 5, 1071–1090.

Cole, D. J., Morgan, B. J. T., and Titterington, D. M. (2010). The parametric structure of models. *Math Biosci* 228, 16–30.

Cowen, L., Besbeas, P. T., Morgan, B. J. T., and Schwarz, C. (2017). Hidden Markov models for extended batch data. *Biometrics* 73, 1321–1331.

deValpine, P. (2012). Frequentist analysis of hierarchical models for population dynamics and demographic data. *J Ornithol* 152, Supplement 2, S393–S408.

Finke, A., King, R., Beskos, A., and Dellaportas, P. (2019). Efficient sequential Monte Carlo algorithms for integrated population models. *J Agric Biol Environ Stat*. https://doi.org/10.1007/s13253-018-00349-9.

Freeman, S. (2009). Towards a method for the estimation and use of averaged multi-species trends, as indicators of patterns of change in butterfly populations. Technical report, Centre for Ecology and Hydrology, Wallingford, OX10 8BB, England.

Gimenez, O., Morgan, B. J. T., and Brooks, S. P. (2009). Weak identiability in models for mark-recapture-recovery data. In Thomson, D. L., Cooch, E. G., and Conroy, M. J., Eds., *Modelling Demographic Processes in Marked Populations*. Springer.

Kéry, M. and Schaub, M. (2012). *Bayesian Population Analysis using WinBUGS: A Hierarchical Perspective*. Academic Press.

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

King, R. (2014). Statistical ecology. *Annu Rev Stat Appl* 1, 401–426.

Knape, J., Jonzén, and Sköld, M. (2011). On observation distributions for state space models of population survey data. *J Anim Ecol* 80, 1269–1277.

Lahoz-Monfort, J. J., Harris, M. P., Wanless, S., Freeman, S. N., and Morgan, B. J. T. (2017). Bringing it all together: Multi-species integrated population modelling of a breeding community. *JABES* 22, 140–160.

Maunder, M. N. and Punt, A. E. (2013). A review of integrated analysis in fisheries stock assessment. *Fish Res* 142, 61–74.

McCrea, R. S. and Morgan, B. J. T. (2014). *Analysis of Capture-recapture Data*. Boca Raton: CRC Press, Chapman & Hall.

Newman, K. B., Buckland, S. T., Morgan, B. J. T., King, R., Borchers, D. L., Cole, D. J., Besbeas, P., Gimenez, O., and Thomas, L. (2014). *Modelling Population Dynamics*. New York: Springer.

Robinson, R. A., Morrison, C. A., and Baillie, S. R. (2014). Integrating demographic data: Towards a framework for monitoring wildlife populations at large spatial scales. *Methods Ecol Evol* 5, 1361–1372.

Schaub, M. and Abadi, F. (2011). Integrated population models: A novel analysis framework for deeper insights into population dynamics. *J Ornithol* 152, 227–237.

Schaub, M. and Fletcher, D. (2015). Estimating immigration using a Bayesian integrated population model: Choices of parametrization and priors. *Environ Ecol Stat* 22, 535–549.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series An Introduction Using R*, Second Edition. Chapman & Hall/CRC.

SUPPORTING INFORMATION

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

How to cite this article: Besbeas P, Morgan BJT. Exact inference for integrated population modelling. *Biometrics*. 2019;75:475–484. https://doi.org/10.1111/biom.13045