

An Integrated Analysis of Multisite Recruitment, Mark-Recapture-Recovery and Multisite Census Data

R.S. Borysiewicz, B.J.T. Morgan, V. Hénau, T. Bregnballe, J.-D. Lebreton, and O. Gimenez

Abstract The statistical analysis of mark-recapture-recovery (MRR) data dates back to the 1960s, when the foundation was laid for stochastic models, fitted to data by the method of maximum likelihood. There have been a number of developments which have proved to be extremely influential. Two of these are: the extension of MRR data and modelling to multi-site inference, and the integrated modelling of single-site MRR and census data. The aim of this study is to unite these two independent research programs, in order to enable effective integrated analysis of multi-site MRR data and multi-site census data. Census data can be described by a state-space model, and the likelihood is formed using the Kalman filter. By making use of movement information provided by MRR data, it is possible to avoid flat likelihood surfaces, thus allowing estimation of site-dependent parameters. This increases the precision of dispersal parameters and allows estimation of parameters inestimable from MRR studies alone.

This paper extends research within the area of integrated population analysis by developing methods for analysing multi-site census data coupled with multi-site capture-recapture data. The methodology is explored using a simulated data set, the structure of which is motivated by a dataset of Great cormorants (*Phalacrocorax carbo sinensis*).

Keywords Integrated Analysis · Kalman Filter · Mark-Recapture-Recovery Data · Multistate Models · *Phalacrocorax carbo sinensis* · Recruitment · State-Space Models

1 Introduction

1.1 Mark-Recapture-Recovery Models

The development of models for mark-recapture-recovery (MRR) data began in the 1960s with the Cormack-Jolly-Seber (CJS) model (Cormack 1964; Jolly 1965;

R.S. Borysiewicz (✉)
IMSAS, University of Kent, Canterbury, UK.
e-mail: R.S. Borysiewicz@kent.ac.uk

Seber 1965), which estimates survival and capture rates from recapture data of an animal population collected at a single site. This model initiated a surge of interest into constructing models for this type of data and numerous extensions have since been developed. These include incorporation of categorical variables for characterising individuals and analysis of multiple data sets using group effects (Lebreton et al. 1992). Also, models that allow the integrated analysis of both recaptures of live animals and recoveries of dead animals have been developed (Burnham 1993; Lebreton et al. 1995; Barker 1997; Catchpole et al. 1998).

A further development which has enhanced the potential of MRR data analysis has been the extension of the CJS model to the multi-site framework. First established by Arnason (1972, 1973) and later developed by Schwarz et al. (1993) and Brownie et al. (1993), this extension allows estimation of survival and transition probabilities as well as recapture parameters and parameter redundancy of these multi-site models has been assessed in Gimenez et al. (2003). The Arnason-Schwarz model has been further generalised to multi-state models (Lebreton et al. 1999) and multi-event models (Pradel 2005). A number of computer software packages for fitting such models are available, including program MARK (White and Burnham 1999) and M-Surge (Choquet et al. 2004).

1.2 Integrated Population Analysis

Integrated population analysis combines data from a variety of sources. Integrating MRR and census data was first proposed by Besbeas et al. (2002) as a method for estimating productivity, otherwise not estimable from either type of data alone. It is possible to form the likelihood for the MRR data as outlined in Section 1.1. A state space model of the census data comprises two parts – the observation equation and the underlying state equation. Gaussian assumptions provide the key to the accessibility of these models and makes it possible to form the likelihood, using a recursive procedure known as the Kalman filter (Harvey 1989; Durbin and Koopman 2001). Assuming independence, it is possible to combine the MRR and census likelihoods to provide one global likelihood which can be optimised to provide maximum likelihood estimates of all parameters.

A multivariate normal approximation to the exact MRR likelihood (Besbeas et al. 2003) more efficiently integrates both data sets. Further, Brooks et al. (2004) introduced a Bayesian approach, while Besbeas et al. (2005) provides a discussion on further possible advances. Besbeas et al. (2008) gives details of the Kalman filter methodology, including initialisation procedures in ecological applications and a discussion of the break-down of independence assumptions and the effects of introducing overdispersion in the state space model.

The integrated analysis performed in this paper is on a simulated multi-site data set which contains both breeding and non-breeding individuals. Recruitment, defined as the progression from a non-breeding state to a breeding state, is a parameter of biological interest, and it is this type of transition, along with dispersals between study areas which will be estimated from the contribution of MRR multi-site/state and multi-site census simulated data.

2 Motivation for t

This investigative simu
corax carbo sinensis, c
Denmark. *P. carbo* is th
in North America, Eur
belong to the Eurasian
the North Atlantic subs
(Hatch et al. 2000).

The data were collec
in 1977, which continu
ulation expansion (198
and recovery data from
who estimated dispersal

The oldest of the six
Ormø (OR, est. 1972)
colonies present in Den
Dyrefod (DY) and Mågø
1985 respectively.

14,018 cormorant ch
metal ring on one leg a
bination of 3 alphanum
cormorants took place fr
only and these breeders
Henaux et al. (2007). R
northern United Kingdo
Romania. Recoveries of
avoid negative bias due t

Each of the six colon
of all occupied nests. Th
were built on the ground
in the other colonies all r

Based on the param
demonstrates the statisti
on multi-site data and al
recruitment structure, ca
framework.

3 Methods

3.1 Formation of the

The closed-form likelih
Brooks (2003). Suppose

2 Motivation for the Study

This investigative simulation study was motivated by data on cormorants, *Phalacrocorax carbo sinensis*, collected by the National Environmental Research Institute in Denmark. *P. carbo* is the most widely distributed of all cormorants – known to breed in North America, Europe, Asia, Africa and Australasia. The studied cormorants belong to the Eurasian subspecies *P. carbo sinensis*. This subspecies is smaller than the North Atlantic subspecies *P. carbo carbo*, and often breeds and winters inland (Hatch et al. 2000).

The data were collected as part of a larger ringing programme started in Denmark in 1977, which continues to the current day. The data correspond to a period of population expansion (1981–1993) in 6 colonies located 32–234 km apart. Recapture and recovery data from this period were analysed in detail by Henaux et al. (2007) who estimated dispersal and recruitment.

The oldest of the six colonies, Vorsø, was established in 1944 and along with Ormø (OR, est. 1972) and Brændegård Sø (BR, est. 1973) comprised the only colonies present in Denmark at the start of the ringing study. Colonies Toft Sø (TO), Dyrefod (DY) and Mågeøerne (MA) established during the study in 1982, 1984 and 1985 respectively.

14,018 cormorant chicks were marked between 1981 and 1991 with a standard metal ring on one leg and a coloured plastic band, engraved with a unique combination of 3 alphanumeric characters, on the other leg. Resightings of the ringed cormorants took place from 1983 to 1993. Resightings were of breeding cormorants only and these breeders were identified using strict biological criteria identified in Henaux et al. (2007). Recoveries spanned a large geographical area, ranging from northern United Kingdom to Southern Algeria and from western Spain to eastern Romania. Recoveries of birds for which only one ring was found were excluded to avoid negative bias due to ring loss.

Each of the six colonies were censused in early May. Data consisted of a count of all occupied nests. The location of nests varies between colonies: in MA nests were built on the ground, in BR nests were found in trees and on the ground while in the other colonies all nests were in trees.

Based on the parameters of this real-life investigation, this simulation study demonstrates the statistical gains of performing integrated population modelling on multi-site data and also the ease with which even complex models, such as the recruitment structure, can be incorporated into an integrated population modelling framework.

3 Methods

3.1 Formation of the Mark-Recapture-Recovery Likelihood

The closed-form likelihood for Arnason-Schwarz models is derived in King and Brooks (2003). Suppose captures or recaptures occur for animals age $j \in \mathcal{J} =$

$\{0, \dots, J\}$ and the study site is split into \mathcal{R} regions. The set of model parameters includes:

$\phi_j(r)$ is the probability that an animal in location r at age j survives until age $j + 1$;

$\lambda_j(r)$ is the probability that an animal in location r at age j dies and is recovered dead before age $j + 1$;

$p_{j+1}(r)$ is the probability that an animal in location r at age $j + 1$ is recaptured; and

$\psi_j(r, s)$ is the probability that an animal in location r at age j moves to location s by age $j + 1$ given that it is alive at age $j + 1$.

The encounter history of each animal can be broken down into three partial histories. These are: last live encounters and beyond, consecutive live sightings and dead recoveries. The likelihood can similarly be deconstructed into these elements and full details of the likelihood construction can be found in the Appendix.

3.2 Formation of the Census Likelihood

3.2.1 The State Space Model and the Kalman Filter

The Kalman filter is a recursive procedure for computing the optimal estimator of a state vector at time t , based on the information available at time t (Harvey 1989). By imposing Gaussian assumptions it is possible to calculate the maximum likelihood estimates of unknown model parameters. The general linear Gaussian state space model is:

$$y_t = Z_t \alpha_t + \epsilon_t \quad (1)$$

$$\alpha_{t+1} = T_t \alpha_t + \eta_t \quad (2)$$

with $\epsilon_t \sim N(0, H_t)$ and $\eta_t \sim N(0, Q_t)$.

Equation (1) is the observation equation and (2) is the state space equation. The state vector, α_t , is unobserved and y_t is a vector of observations. The matrices Z_t , T_t , H_t and Q_t are assumed to be serially independent and independent of each other at all times.

The initial state vector α_1 is assumed to be $N(a_1, P_1)$ independently of $\epsilon_1, \dots, \epsilon_n$ and η_1, \dots, η_n . In practice, some or all of the matrices will depend on elements of an unknown model parameter vector.

The aim is to obtain a conditional distribution of α_{t+1} given $Y_t = \{y_1, \dots, y_t\}$. Since all distributions are normal, conditional distributions of subsets of variables given other subsets are also normal; the required distribution is therefore determined by a knowledge of $a_{t+1} = E(\alpha_{t+1}|Y_t)$ and $P_{t+1} = Var(\alpha_{t+1}|Y_t)$.

The Kalman
the smoothed e
 y_1, \dots, y_n . Den
of normality, the

$L(\theta$

and the log-lik

where $p(y_1|Y_0)$
obtain

log

This likelihood
likelihood since v_t
is the covariance
vectors v_t and m

3.2.2 Ecologica

Define N_{it}^x to be
non-breeder and

then following th

Vector y_t is th
to be constant o
underlying state s
the explicit form

The Kalman f
and covariance m
and Morgan (200

The Kalman Filtering derives the filtering equations and may also compute the smoothed estimates of the error vectors ϵ_t and η_t , given all the observations y_1, \dots, y_n . Denoting the parameter vector by θ and using our previous assumptions of normality, the likelihood is

$$L(\theta | y) = p(y_1, \dots, y_n | \theta) = p(y_1 | \theta) \prod_{t=2}^n p(y_t | Y_{t-1}, \theta)$$

and the log-likelihood is given by

$$\log L(\theta | y) = \sum_{t=1}^n \log p(y_t | Y_{t-1}, \theta)$$

where $p(y_1 | Y_0, \theta) = p(y_1)$. Following substitution of appropriate parameters we obtain

$$\log L(\theta | y) = -\frac{np}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^n \log(|F_t| + v_t' F_t^{-1} v_t) \quad (3)$$

This likelihood is known as the prediction error decomposition form of the likelihood since v_t can be interpreted as a vector of prediction errors, $y_t - E(y_t)$. F_t is the covariance matrix of the conditional distribution of the observations and both vectors v_t and matrix F_t are calculated directly from the Kalman filter.

3.2.2 Ecological Application of the Kalman Filter

Define N_{it}^x to be the number of animals in state $x = \{N, B\}$ (where N denotes non-breeder and B denotes breeder) in site $i = 1, 2, 3$ at time t . State vector α_t is

$$\alpha_t = (N_1^N \ N_2^N \ N_3^N \ N_1^B \ N_2^B \ N_3^B)'_t, \quad (2)$$

then following the notation of equation (1), since only breeders are observed

$$Z_t = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \forall t.$$

Vector y_t is the observed census data and the observation error, σ , is assumed to be constant over time and site; we discuss relaxing this assumption later. The underlying state space equation is governed by a Leslie matrix (Caswell 2000), and the explicit form of the Leslie matrix for this application is given in Section 3.4.

The Kalman filter is initialised by specifying values for a_1 and P_1 , the mean and covariance matrix of the initial state vector, respectively. Following Besbeas and Morgan (2008) we initialised the filter using the stable age distribution of the

population. Other initialisation procedures were implemented on the same simulated data set and the model structure was robust to the assumption of the stable age distribution start.

3.3 Simulation of Multisite Recruitment Data

150 simulated non-breeding animals were marked at each site and MRR data were simulated on this cohort for seven years. All ringing was carried out on non-breeders and recaptures were of breeders only. Recapture and survival probabilities were site- and state-dependent and the recovery probability was independent of recovery site. Three types of transition were considered, which reflect the model structure used by Henaux et al. (2007):

- **natal dispersal:** the movement of a non-breeder from one geographical location to another whilst remaining a non-breeder.
- **recruitment:** the accession from non-breeder to breeder.
- **breeding dispersal:** the movement of a breeder from one geographical location to another.

Cormorants can only start breeding from age two and remain capable of breeding in every subsequent year (Henaux et al. 2007), thus, within this simulation study we assume that animals start recruiting at age 2 and that all animals have recruited by age 5 and once a breeder, they remain a breeder until death. Only breeding birds are ever recaptured and hence constraints applied to natal dispersal and recruitment need to be enforced in order to ensure identifiability of all transitions, and so it is assumed that natal dispersal occurs only in the first year of life and subsequently recruitment then occurs at a single site and no further dispersals occur between sites until the animal has reached the breeding state.

The assumptions, defined above, are in addition to the traditional Arnason-Schwarz model assumptions, and in order to proceed with integrated analysis the MRR data and census data must be independent. Violation of this assumption can lead to biased estimates, as shown in Besbeas et al. (2008).

The simulated census data were site-specific with structure governed by a Leslie matrix. Twenty years of site-dependent census information were simulated which, like the cormorant data set, was of breeding birds only. Parameter values were chosen to reflect biologically reasonable values. Fecundity (defined to be the number of offspring multiplied by the probability of reproduction in a particular year) was fixed at a constant value of 1.2. Fecundity however could be adapted to allow for time, site or density dependence. Transition, survival, capture and recovery probabilities were set to values which were biologically reasonable, with non-breeder survival (0.6) assumed lower than breeder survival (0.8) and recovery rates (0.4) lower than recapture rates (0.7). Dispersal rates varied between 0.05 and 0.3 between sites.

3.4 Specification

Suppose we denote the following states as 1, 2 and 3, and the transitions involved in the MRR

Survival Probabilities:

$r \in \{1, 2, 3\}$ denote survival;

Recovery Probabilities:

Recapture Probabilities:

Natal Dispersal:

Breeding Dispersal:

Recruitment: ψ_{2+}

Non-Maturation:

The structure of the transition matrix above is then:

$$\begin{pmatrix} \phi(4)\psi_{2+}(4, 4) & 0 & 0 \\ 0 & \phi(5)\psi_{2+} & 0 \\ 0 & 0 & 0 \\ \phi(4)\psi_{2+}(4, 1) & 0 & 0 \\ 0 & \phi(5)\psi_{2+} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

where f is the fecundity.

3.5 Computational

MATLAB was used to construct the census likelihoods, assuming two likelihoods, assuming a maximum likelihood estimation method within a multinomial framework, strain recapture, recovery, and a generalised logit link, as well as transition probabilities of the transitions added to the model. The appropriate likelihoods can be found in Choquet et al. (2004) for data.

3.4 Specification of the Leslie Matrix

Suppose we denote the three non-breeding states as 4, 5 and 6 and the three breeding states as 1, 2 and 3 for the respective three sites in each case. The parameters involved in the MRR likelihood, following the notation of Section 3.1, are:

Survival Probability: $\phi_j(r) = \phi(r)$ for all j , for $r \in \{1, 2, 3, 4, 5, 6\}$, where $r \in \{1, 2, 3\}$ denotes breeder survival and $r \in \{4, 5, 6\}$ denotes non-breeder survival;

Recovery Probability: $\lambda_j(r) = \lambda$ for all j and r ;

Recapture Probabilities: $p_{j+1}(r) = p(r)$ for all j and $r \in \{1, 2, 3\}$;

Natal Dispersal: $\psi_1(r, s)$ for $r, s \in \{4, 5, 6\}$

Breeding Dispersal: $\psi(r, s)$ for $r, s \in \{1, 2, 3\}$

Recruitment: $\psi_{2+}(r, r-3)$ for $r \in \{4, 5, 6\}$

Non-Maturation: $\psi_{2+}(r, r)$ for $r \in \{4, 5, 6\}$

The structure of the Leslie matrix for the state space model, using the notation above is then:

$$\begin{pmatrix} \phi(4)\psi_{2+}(4, 4) & 0 & 0 & f\phi(4)\psi_1(4, 4) & f\phi(5)\psi_1(5, 4) & f\phi(6)\psi_1(6, 4) \\ 0 & \phi(5)\psi_{2+}(5, 5) & 0 & f\phi(4)\psi_1(4, 5) & f\phi(5)\psi_1(5, 5) & f\phi(6)\psi_1(6, 5) \\ 0 & 0 & \phi(6)\psi_{2+}(6, 6) & f\phi(4)\psi_1(4, 6) & f\phi(5)\psi_1(5, 6) & f\phi(6)\psi_1(6, 6) \\ \phi(4)\psi_{2+}(4, 1) & 0 & 0 & \phi(1)\psi(1, 1) & \phi(2)\psi(2, 1) & \phi(3)\psi(3, 1) \\ 0 & \phi(5)\psi_{2+}(5, 2) & 0 & \phi(1)\psi(1, 2) & \phi(2)\psi(2, 2) & \phi(3)\psi(3, 2) \\ 0 & 0 & \phi(6)\psi_{2+}(6, 3) & \phi(1)\psi(1, 3) & \phi(2)\psi(2, 3) & \phi(3)\psi(3, 3) \end{pmatrix}$$

where f is the fecundity parameter.

3.5 Computational Implementation

MATLAB was used to code the MRR likelihood and the Kalman filter which constructed the census likelihood. The global likelihood, formed by multiplying the two likelihoods, assuming independence, was then optimised using a built-in optimisation method within the MATLAB software. The logistic link was used to constrain recapture, recovery and survival probabilities between 0 and 1, whilst the generalised logit link, which is an extended logit function was used to ensure that as well as transition probabilities being between 0 and 1, appropriate combinations of the transitions added to 1. Further details of the use of the generalised logit link can be found in Choquet et al. (2005). Figure 1 demonstrates the formation of the appropriate likelihoods and also the common parameters used to model each type of data.

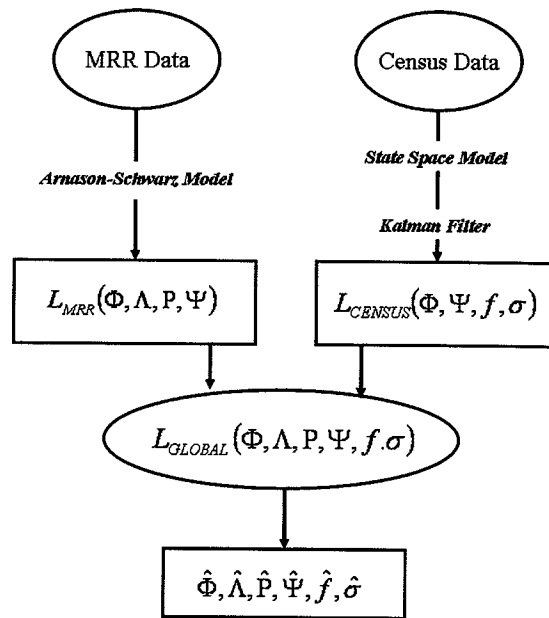


Fig. 1 Directional acyclic graph showing how two types of data (MRR data and census data) are combined to obtain global estimates of parameters: Φ – survival parameters; Λ – recovery parameters; P – recapture parameters; Ψ – transition parameters, f – fecundity, σ – observation error. The Arnason-Schwarz model is used to form the MRR likelihood whilst the Kalman filter is applied to a state space model to form the census likelihood

4 Results

The maximum likelihood estimates (MLEs) are shown in Table 1. There were no issues of either intrinsic or extrinsic parameter redundancy in the integrated model structure, so the MRR data analysed alone is able to estimate all survival, transition and capture/recovery probabilities. Once the MRR data were combined with census information we were also able to estimate the fecundity of the population. The addition of census data improves the precision of capture and recovery probabilities slightly (Table 2), however the change is small as the only information added by the census data is through the correlation structure of these parameters. The largest improvement in precision is for the breeding dispersal parameters, with no improvement in the natal dispersal parameter. This is most likely due to the fact that natal dispersal only appears in the Leslie matrix as a product with the unknown parameter fecundity. Thus, precision is not improved without compromising the precision of the fecundity parameter. Attempting to estimate parameters from the census data alone results in parameter estimates with low precision due to the census data producing a flat likelihood (the fecundity estimate from census data alone is 1.77 with standard error 14.854), however once the census data are combined with MRR data,

Table 1 The maximum likelihood estimates and standard errors of the parameters of the integrated model

Parameter
$\phi(1, 2, 3)$
$\phi(4, 5, 6)$
p
$\psi(4, 5) = \psi(4, 6)$
$\psi(5, 4) = \psi(5, 6)$
$\psi(6, 4) = \psi(6, 5)$
$\psi(1, 2) = \psi(1, 3)$
$\psi(2, 1) = \psi(2, 3)$
$\psi(3, 1) = \psi(3, 2)$
$\psi(4, 1) = \psi(4, 2)$
$\psi(6, 3)$
λ
f

Table 2 The maximum likelihood estimates and standard errors of the parameters of the integrated model

Parameter
$\phi(1, 2, 3)$
$\phi(4, 5, 6)$
p
$\psi(4, 5) = \psi(4, 6)$
$\psi(5, 4) = \psi(5, 6)$
$\psi(6, 4) = \psi(6, 5)$
$\psi(1, 2) = \psi(1, 3)$
$\psi(2, 1) = \psi(2, 3)$
$\psi(3, 1) = \psi(3, 2)$
$\psi(4, 1) = \psi(4, 2)$
λ
f
σ

the fecundity parameter has some precision

The general structure of the integrated model is combining the additional information from the common parameters

The combining of the census data with the MRR data allows obtaining a joint likelihood $N(0, y_{t-1}\sigma^2)$ (T) as density dependent

Table 1 The maximum likelihood estimates and associated standard errors calculated by a finite difference method, of the MRR data used alone and the integrated MRR and census data

Parameter	True Value	MRR Estimate	MRR SE	Integrated Estimate	Integrated SE
$\phi(1, 2, 3)$	0.80	0.826	0.0312	0.820	0.0279
$\phi(4, 5, 6)$	0.60	0.587	0.0211	0.586	0.0211
p	0.70	0.714	0.0396	0.712	0.0379
$\psi(4, 5) = \psi(4, 6)$	0.10	0.078	0.0395	0.073	0.0350
$\psi(5, 4) = \psi(5, 6)$	0.20	0.189	0.0441	0.193	0.0418
$\psi(6, 4) = \psi(6, 5)$	0.05	0.024	0.0227	0.025	0.0236
$\psi(1, 2) = \psi(1, 3)$	0.20	0.248	0.0367	0.209	0.0213
$\psi(2, 1) = \psi(2, 3)$	0.10	0.083	0.0283	0.106	0.0194
$\psi(3, 1) = \psi(3, 2)$	0.30	0.332	0.0351	0.325	0.0279
$\psi(4, 1) = \psi(5, 2) = \psi(6, 3)$	0.60	0.593	0.0593	0.601	0.0559
λ	0.40	0.406	0.0258	0.405	0.0258
f	1.2	-	-	1.167	0.1420

Table 2 The mean square errors of the parameters for the MRR data used alone and the integrated MRR and census data

Parameter	MSE (MRR Data Alone)	MSE (Integrated Data)
$\phi(1, 2, 3)$	0.0752	0.0525
$\phi(4, 5, 6)$	0.0106	0.0110
p	0.0422	0.0375
$\psi(4, 5) = \psi(4, 6)$	0.4436	0.4621
$\psi(5, 4) = \psi(5, 6)$	0.1504	0.1273
$\psi(6, 4) = \psi(6, 5)$	1.6180	1.5683
$\psi(1, 2) = \psi(1, 3)$	0.2402	0.0362
$\psi(2, 1) = \psi(2, 3)$	0.2184	0.0594
$\psi(3, 1) = \psi(3, 2)$	0.1733	0.1056
$\psi(4, 1) = \psi(5, 2) = \psi(6, 3)$	0.0612	0.0544
λ	0.0120	0.0119
f	-	0.0156
σ	-	0.0232

the fecundity parameter estimate was precise (MSE = 0.0156). Adult survival gains some precision whilst juvenile survival does not.

The generalised variances of the common parameters of the MRR analysis and the integrated analysis are 3.64×10^{-13} and 2.80×10^{-17} respectively. Thus, combining the additional census data has considerably improved the overall precision of common parameters and allowed fecundity to be estimated with good precision.

The combined analysis has also accurately estimated the observation error of the census data. Though it was assumed constant for this simulation, it is interesting to allow observation error to vary proportionally to population size, e.g. $\epsilon_t \sim N(0, y_{t-1}\sigma^2)$ (Tavecchia et al. 2006). Similarly, complex fecundity structures, such as density dependence could also be considered.

5 Conclusions

These procedures are extensions of the single-site integrated population analysis. MRR is now a widely used type of data, with even advanced MRR models in common use. Frequently, however, census information is collected concurrently with MRR data. Until now, census data have been analysed separate from the MRR data, thereby sacrificing their shared information. The methodologies described in this paper are simple to implement and can be completed 'post-hoc' by using a simple multivariate normal approximation to the MRR likelihood; if only MLEs and associated standard errors are available, they can be used to construct an approximate diagonal variance-covariance matrix to facilitate analysis and complicated population transitions can be incorporated into the state space model.

The potential of single-site integrated population analysis has been assessed in numerous studies as a method for estimating previously inestimable parameters. This simulation study has shown that not only does multi-site/state integrated population analysis estimate these parameters, but also greatly improves the precision of some parameters. This is obviously desirable in complex models with a large number of parameters that are frequently estimated with low precision.

Single site integrated population modeling with one dimensional census information allows the estimation of additional parameters, however little change is made to the precision of common parameters. By incorporating multi-dimensional census data, in terms of site or state-specific census data, we have shown that this extra information has greatly improved the precision of parameters within the model.

Acknowledgments We wish to thank Jens Gregersen and Lars Abrahamsen for their great effort in ringing and recording breeding attempts of colour-ringed cormorants.

Appendix

The Construction of the Closed-Form Arnason-Schwarz Likelihood

This appendix gives an outline of the construction of the explicit MRR likelihood first derived in King and Brooks (2003). Recall that the encounter history of an animal can be broken down into three partial histories: last live encounters and beyond, consecutive live sightings, dead recoveries. The likelihood can similarly be deconstructed into these elements. The following derived probabilities facilitate the likelihood construction.

Let $\chi_{(j,k)}(r)$ denote the probability that an animal is seen for the last time at age $j \in \mathcal{J}$ in location $r \in \mathcal{R}$, and would be age k at the end of the study, with $j \leq k \leq \mathcal{J}$. Then,

$$\chi_{(j,k)}(r) = \begin{cases} 1 & (j = k) \\ 1 - \phi_{(j)}(r) \left[1 - \sum_{s \in \mathcal{R}} \psi_j(r, s) \{1 - p_{j+1}(s)\} \chi_{(j+1,k)}(s) \right] & (j < k) \\ -\{1 - \phi_j(r)\} \lambda_j(r) & \end{cases}$$

An Integrat

$O_{(k,j)}(r)$,
remains u
for $0 \leq k$

where $Q_{(j,k)}$
at age k
unobserv

$Q_{(k,j)}$

$D_{(k,j)}(r)$ c
($j, j + 1$)

$D_{k,j}$

The foll
data are the

$v_{(j,k)}(r)$
loca
 $n_{(k,j)}(r)$
at ag
 $d_{(k,j)}(r)$
 $j +$

The likel

$L_{(j,k)}$

where θ cor
likelihood by

$O_{(k,j)}(r, s)$ denotes the probability that an animal in location $r \in \mathcal{R}$ at age $k \in \mathcal{J}$ remains unobserved until it is subsequently resighted in location $s \in \mathcal{R}$ at age $j + 1$, for $0 \leq k \leq j \leq J - 1$. Then,

$$O_{(k,j)}(r, s) = p_{j+1}(s)Q_{(k,j)}(r, s),$$

where $Q_{(k,j)}(r, s)$ denotes the probability that an animal migrates from region $r \in \mathcal{R}$ at age $k \in \mathcal{J}$ to location $s \in \mathcal{R}$ at age $j + 1$, for $0 \leq k \leq j \leq J - 1$, and is unobserved between these ages, and is given by

$$Q_{(k,j)}(r, s) = \begin{cases} \phi_k(r)\psi_k(r, s) & (k = j) \\ \phi_k(r) \sum_{l=1}^R \{1 - p_{k+1}(l)\}\psi_k(r, l)Q_{k+1,j}(l, s) & (k < j) \end{cases}$$

$D_{(k,j)}(r)$ denotes the probability that an animal is recovered dead in the interval $(j, j + 1)$ given that it is last seen at age $k \leq j$ in $r \in \mathcal{R}$ and is given by

$$D_{kj}(r) = \begin{cases} \{1 - \phi_j(r)\}\lambda_k(r) & (k = j) \\ \sum_{l=1}^R \{1 - \phi_j(l)\}\lambda_j(l)\{1 - p_j(l)\}Q_{k,j-1}(r, l) & (k < j) \end{cases}$$

The following sufficient statistics which are obtained from the encounter history data are then formed:

$v_{(j,k)}(r)$ denotes the number of animals that are recaptured for the last time in location $r \in \mathcal{R}$ aged j and would be aged $j \leq k \leq J$ at the end of the study; $n_{(k,j)}(r, s)$ denotes the number of animals that are observed in location $r \in \mathcal{R}$ at age k and next observed alive in location $s \in \mathcal{R}$ at age $j + 1$; and $d_{(k,j)}(r)$ denotes the number of animals recovered dead between ages j and $j + 1$ that are last observed alive at age $k \leq j$ in location $r \in \mathcal{R}$.

The likelihood function has the form given below:

$$L(\theta|v, n, d) = \prod_{r \in \mathcal{R}} \left[\prod_{j=0}^J \prod_{k=j}^J \{\chi_{(j,k)}(r)\}^{v_{(j,k)}(r)} \prod_{k=0}^{J-1} \prod_{j=k}^{J-1} \{D_{(k,j)}\}^{d_{(k,j)}(r)} \prod_{k=0}^{J-1} \prod_{j=k}^{J-1} \prod_{s \in \mathcal{R}} \{O_{(k,j)}(r, s)\}^{n_{(k,j)}(r,s)} \right]$$

where θ comprises the model parameters $\{\Phi, \Lambda, P, \Psi\}$ and we denote our MRR likelihood by $L_{MRR}(\Phi, \Lambda, P, \Psi)$.

References

- Arnason AN (1972) Parameters estimates from mark-recapture-recovery experiments on two populations subject to migration and death. *Researches on Population Ecology* 13:97–113.
- Arnason AN (1973) The estimation of population size, migration rates, and survival in a stratified population. *Researches on Population Ecology* 15:1–8.
- Barker RJ (1997) Joint modeling of live-recapture, tag-resight, and tag-recovery data *Biometrics* 53(2):666–677.
- Besbeas P, Morgan BJT (2008) Kalman filter initialization for modelling population dynamics. *Biometrics Invited Revision*.
- Besbeas P, Freeman SN, Morgan BJT, Catchpole EA (2002) Integrating mark-recapture-recovery and census data to estimate animal abundance and demographic parameters. *Biometrics* 58: 540–547.
- Besbeas P, Lebreton J-D, Morgan BJT (2003) The efficient integration of abundance and demographic data. *Applied Statistics* 52:95–102.
- Besbeas P, Freeman SN, Morgan BJT (2005) The potential of integrated population modelling. *Australian & New Zealand Journal of Statistics* 47(1):35–48.
- Besbeas P, Borysiewicz RS, Morgan BJT (2008) Completing the ecological jigsaw. In press.
- Brooks SP, King R, Morgan BJT (2004) A Bayesian approach to combining animal abundance and demographic data. *Animal Biodiversity and Conservation* 27(1):515–529.
- Brownie C, Hines JE, Nichols JD, Pollock KH, Hestbeck JB (1993) Capture-recapture studies for multiple strata including non-markovian transitions. *Biometrics* 49:1173–1187.
- Burnham KP (1993) A theory for combined analysis of ring-recovery and recapture data. pages 199–213. Basel: Birkhauser Verlag.
- Caswell H (2000) *Matrix population models: Construction, analysis and interpretation*. Sinauer Associates Incorporated, Sunderland, MA.
- Catchpole EA, Freeman SN, Morgan BJT, Harris MP (1998) Integrated recovery/recapture data analysis. *Biometrics* 54:33–46.
- Choquet R, Reboulet A-M, Pradel R, Gimenez O, Lebreton J-D (2004) M-Surge: New software specifically designed for multistate capture-recapture models. *Animal Biodiversity and Conservation* 27(1):207–215.
- Choquet R, Reboulet A-M, Pradel R, Gimenez O, Lebreton J-D (2005) M-Surge 1–7 user's manual. CEFE/CNRS, Montpellier, FRANCE, http://ftp.cefe.cnrs.fr/biom/Soft_CR.
- Cormack RM (1964) Estimates of survival from sightings of marked animals. *Biometrika* 51: 429–438.
- Durbin J, Koopman SJ (2001) *Time series analysis by state space methods*. Oxford University Press, Oxford, U.S.
- Gimenez O, Choquet R, Lebreton J-D (2003) Parameter redundancy in multistate capture-recapture models. *Biometrical Journal* 6:704–722.
- Harvey AC (1989) *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, Cambridge, U.K.
- Hatch JJ, Brown KM, Hogan GG, Morris RD (2000) Great cormorant (*Phalacrocorax carbo*). In: *The birds of North America*, Eds. A. Poole and F.Gill, (553).
- Henaus V, Bregnballe T, Lebreton J-D (2007) Dispersal and recruitment during population growth in a colonial bird, the great cormorant. *Journal of Avian Biology* 38(1):44–57.
- Jolly GM (1965) Explicit estimates from capture-recapture data with both death and immigration-stochastic models. *Biometrika* 52:225–247.
- King R, Brooks SP (2003) Closed-form likelihoods for Arnason-Schwarz models. *Biometrika* 90(2):435–444.
- Lebreton J-D, Burnham KP, Clobert J, Anderson DR (1992) Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs* 62:67–118.
- Lebreton J-D, Morgan BJT, Pradel R, Freeman SN (1995) A simultaneous analysis of dead recovery and live recapture data. *Biometrics* 51(4):1418–1428.

- Lebreton J-D, Almeras T, Pradel R (1999) Competing events, mixtures of information and multistratum recapture models. *Bird Study*, 46(Suppl.):S39-46.
- Pradel R (2005) Multievent: An extension of multistate capture-recapture models to uncertain states. *Biometrics* 61.
- Schwarz CG, Schweigert JF, Arnason AN (1993) Estimating migration rates using tag-recovery data. *Biometrics* 59:291-318.
- Seber GAF (1965) A note on the multiple-recapture census. *Biometrika* 52:249-259.
- Tavecchia G, Besbeas PT, Coulson T, Clutton-Brock TH, Morgan BJT (2006) Estimating population size and hidden demographic parameters with state-space modelling. Technical Report UKC/IMS/06/15.
- White GC, Burnham KP (1999) Program Mark: Survival estimation from populations of marked animals. *Bird Study* 46:S120-S138.