

SPECIAL ISSUE. MODELLING DEMOGRAPHIC PROCESSES IN MARKED POPULATIONS: PROCEEDINGS OF THE EURING 2013 ANALYTICAL MEETING

Spatially explicit integrated population models

Richard B. Chandler^{1*} and Joseph D. Clark²

¹Warnell School of Forestry and Natural Resources, University of Georgia, Athens, GA 30602, USA; and ²USGS Southern Appalachian Research Branch, Knoxville, TN 37901, USA

Summary

1. Studies of demographic processes are typically restricted to small geographic areas and short time periods due to the costs of marking and monitoring individuals. However, environmental changes are occurring at much broader spatial and temporal scales, and thus, inferences about the mechanisms governing population dynamics need to be scaled accordingly. Recently developed integrated population models (IPMs) represent an approach for doing so, by jointly analysing survey data and capture–recapture data.

2. Although promising, several shortcomings of conventional IPMs exist, including difficulties accounting for spatial variation in demographic, movement and detection parameters; limited ability to make spatially explicit predictions of abundance or vital rates; and a requirement that the survey data and the capture–recapture data are independent. We demonstrate how each of these limitations can be resolved by adopting a spatial population dynamics model upon which both the survey data and the capture–recapture data are conditioned.

3. We applied the model to 6 years of hair data collected on the threatened Louisiana black bear *Ursus americanus luteolus*. For years in which the hair samples were genotyped, the resulting data are information-rich (but expensive) spatial capture–recapture (SCR) data. For the remaining years, the data are binary detection data, of the type often analysed using occupancy models. We compared estimates of demographic parameters and annual abundance using various combinations of the SCR and detection data, and found that combining the SCR data and the detection data resulted in more precise estimates of abundance relative to estimates that did not use the detection data. A simulation study provided additional evidence of increased precision, as well as evidence that the estimators of annual abundance are approximately unbiased.

4. The ability to combine survey data and capture–recapture data using a spatially explicit model opens many possibilities for designing cost effective studies and scaling up inferences about the demographic processes influencing spatial and temporal population dynamics.

Key-words: count data, detection data, integrated population models, Louisiana black bear, population dynamics, spatial capture–recapture, spatial scaling

Introduction

A central challenge in applied population ecology is to understand the factors affecting population dynamics at large spatial and temporal scales (Post & Forchhammer 2002; Clark *et al.* 2004; Melbourne & Chesson 2005). Such information is needed to describe and predict the population-level consequences of processes such as global climate and land use change (Mazerolle *et al.* 2005; Grosbois *et al.* 2008). However, a mechanistic description of population dynamics requires knowledge about demographic processes (Clutton-Brock & Sheldon 2010), and estimating demographic parameters typically requires arduous and expensive capture–recapture studies in which individuals are marked and followed through time (Williams, Nichols & Conroy 2002; Royle *et al.* 2013).

Due to the costs associated with capture–recapture methods, most demographic studies are conducted at small spatial scales

and often over relatively short time periods. To study population processes at larger spatial and temporal extents, ecologists often resort to simple count-based or presence–absence studies, which are relatively inexpensive to implement. Although such survey data can be useful for studying trends in abundance or patch dynamics (Chandler & King 2011; Mackenzie *et al.*, 2006), they contain little information about the demographic processes governing population dynamics (Brooks, King & Morgan 2004).

Recently developed integrated population models (IPMs, Besbeas, Lebreton & Morgan 2002 2003; Brooks, King & Morgan 2004; Schaub *et al.* 2007) represent a powerful approach for combining capture–recapture data and survey data to make inferences about the demographic processes affecting population dynamics at large spatial and temporal extents. In their most basic form, a joint analysis is conducted in which abundance is estimated by fitting a state-space model to count data, and demographic parameters are estimated by fitting a capture–recapture model to the data on marked individu-

*Correspondence author. E-mail: rchandler@warnell.uga.edu

als. State-space models (de Valpine & Hastings 2002; Buckland *et al.* 2004) are time-series models describing temporal changes in abundance while accounting for process variation (e.g. demographic and environmental stochasticity) and observation error. Often, classical population growth models (e.g. exponential or logistic) are used to model population growth, and the process variation is assumed to be normally distributed. The purpose of combining state-space models with capture–recapture data in an IPM is to allow for a mechanistic description of population dynamics that includes mortality, recruitment and movement parameters, rather than the less informative summary parameters (e.g. population growth rate) comprising standard state-space models (Schaub & Abadi 2011).

Although IPMs offer tremendous promise for scaling up studies of population dynamics, development of these models is still in its infancy, and conventional models share several important limitations. For example, survival, productivity and even detection probability are best thought of as individual-level processes; yet, most IPMs do not fully accommodate individual heterogeneity, thereby restricting their utility and possibly resulting in biased parameter estimates (Abadi, Botha & Altwegg 2013). In addition, the process error and observation error used in the state-space models are typically modelled as Gaussian random variables, even though abundance and count data are integer-valued. If population size is large, the Gaussian assumption may be inconsequential, but it can lead to unreasonable predictions (e.g. negative values of abundance) if population size is small (Brooks, King & Morgan 2004). Another limitation of conventional IPMs is that they assume the capture–recapture data, and the survey data are independent of one another such that the joint likelihood can be computed easily (but see Mazzetta, Morgan & Coulson 2010). This assumption will often be violated in the common situation in which both data sets are collected in the same region. Parameter estimation poses yet another problem because many state-space models require unrealistic assumptions about abundance in the initial time period (e.g. that the population is in equilibrium) in order for the parameters to be identifiable (Polansky *et al.* 2009). Finally, the fact that conventional IPMs are not spatially explicit poses its own problems. For instance, it can be difficult to infer abundance in unsampled regions when using non-spatial models (Best, Ickstadt & Wolpert 2000), and it is difficult to reconcile the effects of scale-dependent processes such as movement when using non-spatial models (Wiens 1989).

The purpose of this paper is to demonstrate how each of these limitations can be overcome using a spatially explicit IPM. The foundation of the model is a spatial point process describing (1) the number and locations of individuals in the population during the initial time period and (2) the variation in abundance (and density) over time as a function of mortality, recruitment and movement. Both the capture–recapture data and the survey data (either count or detection/non-detection data) can be modelled conditional on the point process using ideas from recently developed spatial capture–recapture (SCR) models (Efford 2004; Borchers & Efford 2008; Gardner *et al.* 2010; Chandler & Royle 2013; Royle *et al.* 2013).

The data

The model described in the next section requires spatially referenced capture–recapture data and survey data. Although these data could arise from many different sampling protocols, we proceed by assuming that the capture–recapture data $\{y_{i,j,k,t}\}$ are binary records indicating if individual i ($i = 1, \dots, I$) was captured at sampling location (e.g. trap) j ($j = 1, \dots, J$) on secondary sampling occasion k ($k = 1, \dots, K$), during primary sampling period t ($t = 1, \dots, T$). This data structure resembles ‘robust design’ capture–recapture data (Pollock 1982), except that the data we consider are trap-specific rather than just ‘trap array’ specific. The Cartesian coordinates of the j th trap are denoted by \mathbf{x}_j . As with the traditional robust design, we assume that there are multiple primary and secondary sampling occasions, and that abundance is constant within, but not among, primary occasions.

The survey data $\{n_{j,k,t}\}$ are also collected during multiple primary and secondary occasions, and they could be either counts or binary detection/non-detection observations (hereafter ‘detection data’). The survey data could be collected at the same set of locations where the capture–recapture sampling occurs, but during different primary periods, or they could be collected during the same primary periods but at different locations from where the capture–recapture data are collected. Alternatively, a combined approach could be used in which the capture–recapture data are collected in a subset of sites during a subset of primary occasions to maximize both the spatial and temporal extents of the study.

The model

The model contains four conditionally related submodels for: (1) initial abundance and distribution; (2) change in abundance and distribution over time, as a function of survival, fecundity and movement parameters; (3) the capture–recapture data; and (4) the survey data. The first two components comprise the state model – the description of the ecological processes of interest – and the second two components are observation models describing the sampling processes.

For inference, we favour a Bayesian approach and Markov chain Monte Carlo (MCMC) because the model has many random variables and constraints, which preclude conventional likelihood analysis. In addition, IPMs inherently involve combining multiple sources of information, which is easily accomplished in a Bayesian analysis.

INITIAL ABUNDANCE AND DISTRIBUTION

We model the initial state of the population (at time $t = 1$) using a spatial point process that describes the initial number of individuals in the population (N_1), as well as their locations ($\mathbf{s}_{i,1}, \dots, \mathbf{s}_{N_1,1}$). In some cases, the coordinates $\mathbf{s}_{i,1}$ might be the exact location of an individual at some point in time, but when modelling capture–recapture data, we define $\mathbf{s}_{i,1}$ as the *average* location of individual i during the first primary period. Otherwise, an additional model is needed to describe movement

within primary periods, which is not of direct interest in this context. We will refer to $\mathbf{s}_{i,1}$ as an ‘activity centre’, although in some contexts, it might have a more specific interpretation such as the centroid of a home range or territory (Efford 2004).

The key element of a spatial point process is the intensity function $\lambda(\mathbf{s})$, which, in this case, describes the expected density of activity centres at location \mathbf{s} in some two-dimensional spatial region \mathcal{S} . If density is constant throughout the region, the point process is said to be homogeneous and the individuals’ locations are uniformly distributed, $\mathbf{s}_{i,1} \sim \text{Uniform}(\mathcal{S})$.

The uniformity assumption might be violated if density varies in response to environmental covariates, a process that could be modelled using a simple loglinear function:

$$\log(\lambda(\mathbf{s})) = \beta_0 + \beta_1 C(\mathbf{s}) \quad \text{eqn 1}$$

where $C(\mathbf{s})$ is a covariate defined at all locations $\mathbf{s} \in \mathcal{S}$. Having specified the intensity function, the expected population size (number of activity centres in \mathcal{S}) at $t = 1$ is given by:

$$\mathbb{E}(N_1) = \Lambda = \int_{\mathcal{S}} \lambda(\mathbf{s}) d\mathbf{s}. \quad \text{eqn 2}$$

Under the assumption $N_1 \sim \text{Poisson}(\Lambda)$, the model is known as an inhomogeneous Poisson point process. The Poisson assumption, however, causes technical difficulties when using MCMC because the dimensions of the parameter space can change with each posterior draw. To resolve this, we adopt what is now a standard data augmentation scheme in which initial abundance is modelled as $N_1 \sim \text{Binomial}(M, \psi)$ where $M \gg I$ is some large integer and $\psi = \Lambda/M$ (Royle, Dorazio & Link 2007; Royle & Dorazio 2012). In general, M should be chosen large enough such that the probability that M individuals were alive during the study, and in \mathcal{S} , is negligible. To facilitate analysis by MCMC and to allow for individual-level covariates, we express the binomial prior on N_1 as a series of Bernoulli outcomes, which identify whether each member of the M individuals is a member of the population and alive ($z_{i,t} = 1$) or not ($z_{i,t} = 0$). Specifically, we have $z_{i,1} \sim \text{Bernoulli}(\psi)$ and $N_1 = \sum_{i=1}^M z_{i,1}$.

SPATIAL AND TEMPORAL CHANGE IN ABUNDANCE AND DISTRIBUTION

The second component of the state model describes abundance and distribution in the years $t = 2, \dots, T$ as a function of survival probability (ϕ) and per capita recruitment (γ). Assuming no permanent emigration, the number of individuals surviving from one time period to the next is $S_t \sim \text{Binomial}(N_{t-1}, \phi)$. If permanent emigration occurs, ϕ could be interpreted as ‘apparent survival’, or the emigration process could be modelled explicitly as demonstrated by Schaub & Royle (2014).

The number of individuals recruited into the population (G_t) can be modelled in a variety of ways. If no information about the sex of individuals in the sample is available, $G_t \sim \text{Poisson}(N_{t-1}\gamma)$ could be used to model density-independent growth. Alternatively, if sex data are available N_{t-1} could be replaced with the number of females in the population. Density-dependent growth could also be considered by

modifying this recruitment model; however, none of these parameterizations would allow for individual variation in survival or recruitment.

A more flexible method of modelling survival and recruitment that can be implemented using MCMC and data augmentation is

$$z_{i,t} \sim \text{Bernoulli}(z_{i,t-1}\phi + (1 - a_{i,t-1})\delta) \quad \text{eqn 3}$$

where $a_{i,t-1} = \max(z_{i,1}, \dots, z_{i,t-1})$ indicates whether the individual is available to be recruited into the population, and δ is the probability that the individual is actually recruited. Equation 3 states that the probability that an individual is alive at time t depends upon survival probability if it was alive at time $t-1$, or upon the recruitment rate if it had never been alive previously. The relationship between the individual-level probability of being recruited (δ) and the per capita recruitment rate (γ) is $\delta = N_{t-1}\gamma/A_{t-1}$, where $A_{t-1} = M - \sum_{i=1}^M a_{i,t-1}$ is the number of individuals available to be recruited in the previous time period. Note that $N_{t-1}\gamma$ could be replaced with any of the recruitment models discussed previously.

Formulating the model in terms of individuals makes it possible to accommodate individual- and time-specific covariates of survival or recruitment, for example, by modelling ϕ and δ as logit-linear functions or as random effects. This formulation of the model also makes it possible to compute population size in each year as a derived parameter $N_t = \sum_{i=1}^M z_{i,t}$ (Gardner *et al.* 2010).

In addition to allowing for individual variation in demographic parameters, the model can also be used to study movement. For example, natal dispersal could be modelled by assuming that the activity centres of adults do not change over time, but the location of a newly recruited individual is governed by a dispersal kernel (e.g. the bivariate normal) centred on the location of the adult’s activity centre. When using data augmentation, this could be implemented as:

$$\mathbf{s}_{i,t} = \begin{cases} \text{BVN}(\mathbf{s}_{w_{i,t},t-1}, \mathbf{\Sigma}) & \text{if } a_{i,t-1} = 0 \\ \mathbf{s}_{i,t-1} & \text{if } a_{i,t-1} = 1 \end{cases} \quad \text{eqn 4}$$

where w indicates which parent individual i was born to. If each adult is equally likely to produce an offspring, the model would be $w_{i,t} \sim \text{Categorical}(\{z_{1,t}/N_t, \dots, z_{M,t}/N_t\})$.

OBSERVATION MODEL FOR THE CAPTURE–RECAPTURE DATA

Capture probability is intuitively a function of the distance between an individual’s home range (which we index by \mathbf{s}) and a trap at \mathbf{x} . Several possible models for distance-related heterogeneity in capture probability are available, but here, we consider the model:

$$y_{ijk} \sim \text{Bernoulli}(p_{ijk})$$

$$p_{ijk} = p_0 \exp(-\|\mathbf{x}_j - \mathbf{s}_i\|^2 / (2\sigma^2))$$

where Euclidean distance is represented as $\|\mathbf{x}_j - \mathbf{s}_i\|$, p_0 is the baseline capture probability when distance is zero, and σ is the scale parameter determining how rapidly capture probability

decreases with distance. This Bernoulli model assumes that capture events are independent of one another, implying that an individual can be captured at multiple traps during a single occasion (as might be the case in hair sampling studies or camera trapping studies). This Bernoulli model is also referred to as the 'proximity detector' model (Efford, Dawson & Borchers 2009). Alternative observation models, such as multinomial or Poisson, could be used if these conditions are not met. As in conventional capture–recapture models, the capture probability parameters can be modelled as functions of individual-specific covariates, time-specific covariates, behavioural effects or random variables (Royle *et al.* 2013).

OBSERVATION MODEL FOR THE SURVEY DATA

Like the model for the capture–recapture data, the model for the survey data is conditional on the underlying spatial population dynamics model. A vast number of possible survey protocols could be considered, but here, we restrict our attention to two possibilities. In the first, the survey data and the capture–recapture data are recorded during different primary periods, but at the same set of locations by the same devices. For example, in the bear study described below, the data were obtained by collecting hair samples from barbed-wire sampling devices (often referred to as 'hair snares' or 'hair snags'). In years when the hair samples are genotyped, the data are spatial capture–recapture (SCR) data, but when the expensive genetic analyses are not conducted, the data are simple binary data indicating whether at least one bear was detected at the sampling location. Chandler & Royle 2013 devised a model for this type of data that retains the underlying capture process described in the previous section. The advantage of this is that it maintains the mechanistic description of how the data arise, and it provides flexibility in designing studies including both marked and unmarked populations (Sollmann *et al.* 2012).

If the sampling device records count data, the observation model is

$$n_{j,k,t} = \sum_{i=1}^N y_{i,j,k,t}. \quad \text{eqn 5}$$

If the sampling method yields binary detection data instead of counts, as in the bear study, an appropriate model is

$$n_{j,k,t} = I\left(\sum_{i=1}^N y_{i,j,k,t} > 0\right) \quad \text{eqn 6}$$

where $I()$ is the indicator function returning 1 if the argument is true and 0 otherwise. The key point is that, regardless of whether the data are counts or binary, they are modelled conditional on the latent capture data. This avoids the requirement of conventional IPMs that the survey data and the capture–recapture data be independent.

A second possible sampling design involves collecting the survey data using an entirely different protocol than that used to collect the capture–recapture data. For example, mist nets might be used at the smaller scale to study avian demographics, while point counts might be used at the larger scale to estimate

population size. In this case, it might be useful to consider the actual location of an individual during each sampling occasion, denoted by \mathbf{u} , rather than just their average location (\mathbf{s}). The locations $\{\mathbf{u}\}$ could be modelled using explicit movement models, or perhaps as a simple function of \mathbf{s} . Once a model for \mathbf{u} is chosen, the number of individuals in plot j during occasions k and t is $N(\mathcal{B})_{j,k,t} = \sum_i I(\mathbf{u}_{i,j,k,t} \in \mathcal{B}_j)$, where $\mathcal{B}_j \in \mathcal{S}$ is the spatial region defining sampling location j . This might be a quadrat or some other fixed-area survey plot. The count data can be modelled conditional on $N(\mathcal{B})$ using any of the standard wildlife sampling protocols such as distance sampling, removal sampling or repeated counts. For example, in the latter case, a reasonable model for the counts is $n_{j,k,t} \sim \text{Binomial}(N(\mathcal{B})_{j,k,t}, p')$ where p' is the detection probability (Royle 2004).

Louisiana black bear study

FIELD METHODS

The Louisiana black bear (*Ursus americanus luteolus*) is one of the 16 recognized subspecies and is listed as threatened under the Endangered Species Act (U.S. Fish and Wildlife Service, 1992). As part of the recovery plan, a monitoring programme was developed to estimate population density and the demographic contributions to population growth (U.S. Fish and Wildlife Service, 1995). Sampling was conducted annually (the primary sampling occasions) at 115 sampling sites, which were single strands of barbed wire surrounding bait (pastries) and a scent lure. Some sites were damaged and thus relocated short distances, resulting in a total of 128 locations being surveyed over the course of the study from 2007 to 2012. The sampling locations are shown in Fig. 1. Each year, hair samples were collected on eight consecutive weeks (the secondary sampling occasions). Due to budget limitations, ≤ 38 samples from each week were genotyped to determine identity and sex, and for the purposes of this paper, we restricted our analysis to this subset of samples. Effort was similar among years, with the exception of 2011 in which a massive flood prevented visits to many hair sampling sites (Table 1). Additional information about the field and genetic methods are given in the Appendix 1.

Although the genetic capture–recapture data are very information-rich, the cost of genotyping the samples is exorbitant, and management agencies are seeking less expensive monitoring methods. One option, described in the previous section, is to forgo the expensive genetic analyses in some years and simply model the detection. In this way, the temporal extent of the study could be increased without substantial increases in cost. Figure 1 illustrates the two types of data that can be obtained under such a design. The two types of data are linked according to Eq. 6, which indicates that some information about population dynamics is contained in the simple detection data. Occupancy models (Mackenzie *et al.*, 2006) are one approach for inference about population dynamics from this type of data, but the resulting inferences are about occurrence dynamics, rather than about abundance and demographics, the focus of this paper.

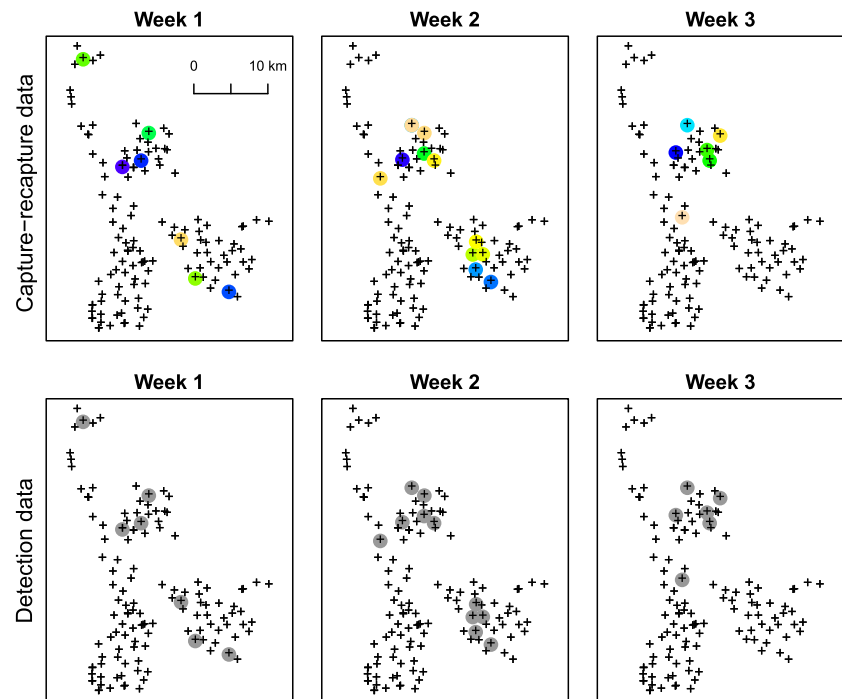


Fig. 1. Two views of the black bear data from the first three sampling occasions in 2007. Crosses indicate the locations of the hair sampling sites. The upper panel shows the capture–recapture data with each bear given a unique colour. The lower panel shows the data coded as simple detection data. The full data set contains eight secondary sampling occasions in each of 6 years.

Table 1. Summary statistics for Louisiana black bear data

Year	2007	2008	2009	2010	2011	2012
Sites sampled	115	115	115	115	51	115
Site visits	919	918	920	918	408	919
Detections	118	227	265	250	226	249
Bears captured	34	47	60	50	29	58

MODEL EVALUATION

To assess the potential of combining spatial capture–recapture and simple detection data, and to determine how much information about population dynamics is contained in the detection data, we fit the most basic form of our model (with no covariates and no dispersal) to the entire 6 years of SCR data as well as to two subsets of the data. The first subset included SCR data from years 2007, 2009 and 2011 but no data from 2008, 2010 and 2012. The second subset included SCR data from 2007, 2009 and 2011, but it also included the detection data from 2008, 2010 and 2012. From here on, we will refer to these as scenarios **S1**, **S2** and **S3**, which are summarized as follows:

S1: All 6 years of SCR data and no detection data

S2: SCR data in 2007, 2009 and 2011 and no data in 2008, 2010 and 2012.

S3: SCR data in 2007, 2009 and 2011, and detection data in 2008, 2010 and 2012.

For each scenario, we generated 1 Markov chain of 8000 posterior samples obtained after discarding 500 adaptive phase samples and 500 burn-in samples. We used the results from the first scenario as the benchmark against which we compared the results from the other two scenarios because it made the most use of the data. We did not consider the potentially important

effects of baiting, year, sex and habitat on parameters because we wished to focus on our primary objective and because a detailed examination of those issues will be the subject of a subsequent paper.

In each scenario, we defined the state-space (S) as a rectangle with lower left and upper right UTM coordinates (615 000, 3 380 000) and (645 000, 3 420 000), respectively. The following vague priors were used in all models: $s \sim \text{Unif}(S)$, $\psi \sim \text{Unif}(0,1)$, $\phi \sim \text{Unif}(0,1)$, $\gamma \sim \text{Unif}(0,3)$ and $\sigma \sim \text{Unif}(0,50\,000)$. MCMC sampling was conducted using JAGS (Plummer 2003), which has two samplers, `dsun` and `dinterval`, that facilitate the implementation of Eqs 5 and 6. JAGS code is provided in Appendix 2.

Results

A total of 109 bears were captured 1356 times. Most individuals were captured multiple times, at multiple sites and in multiple years (Fig. 2, Table 1). Such characteristics are ideal for fitting SCR models because the precision of the estimate of σ depends on the number of spatial recaptures. Nonetheless, some of the variation in the data that is evident in Fig. 2 was unlikely to have been well described by our simplistic model. For instance, one individual was captured 129 times, more than twice as often as any other bear. This might have been due to ‘trap happiness’ or several other factors that we could have modelled, but did not in the interests of focusing on the primary objectives of the paper.

Posterior summaries of annual abundance are presented in Fig. 3. The posteriors appeared to have converged based on the Heidelberg & Welch 1983 diagnostic and visual inspections. Not surprisingly, the most precise estimates were obtained under **S1**, the scenario in which all 6 years of SCR

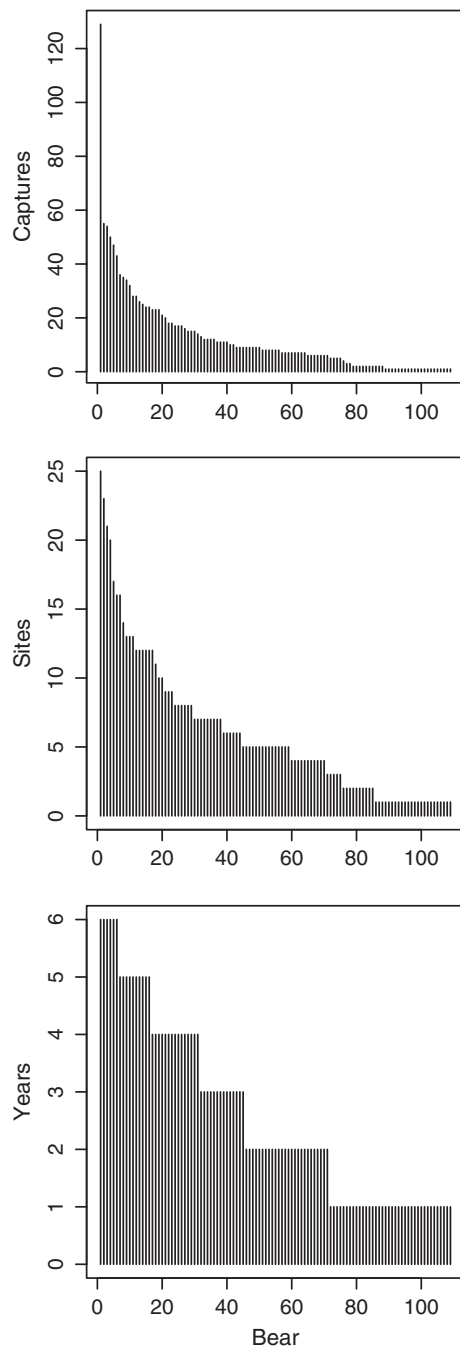


Fig. 2. The number of times each bear was 'captured' (top), the number of sites of capture for each bear (middle) and the number of years in which each bear was captured (bottom). Each vertical line is the frequency for a particular bear.

data were utilized. Precision was lower under scenario 2, in which SCR data were only used in 2007, 2009 and 2011. As shown in Table 2, the posterior variances for yearly abundance were 3–565% higher under **S2** than under **S1**, reflecting the fact that half of the data were not used. For the third scenario in which 3 years of SCR data were combined with 3 years of binary detection data, posterior precision was intermediate between **S1** and **S2** (Table 2), indicating that the binary detection data do contain information about population dynamics.

The posterior variances for **S3** were 0.7–117% higher than for **S1**.

Although posterior precision was substantially improved by including the binary detection data, we could not assess accuracy using these data because the true parameters were unknown. However, we compared posterior mean point estimates from **S2** and **S3** to estimates from **S1** as an index of model performance. For both **S2** and **S3**, the posterior means were on the order of 3–5% lower during the first 3 years than the point estimates from **S1** (Table 2). For the latter 3 years, the posterior means for **S2** were 1.7–9.5% higher than those of **S1**. In contrast, for **S3**, the posterior means were 7–14% lower. These differences reflect patterns in the posterior distributions of apparent survival (Fig. 4). Specifically, the posterior mean of ϕ was 0.826 for **S1**, compared with the higher estimate of 0.844 for **S2**, and the lower estimate of 0.782 for **S3**. In contrast, the posterior distributions of γ were similar for all three scenarios, although the variance was 23% higher for **S2** than for **S3**.

Simulation study

SETTINGS AND MODEL FITTING

We conducted a simulation study to determine whether the results from the black bear analysis were generalizable, and to assess bias, which was not possible using the bear data. We generated 100 data sets from the model with the following parameters: initial abundance $E[N_1] = M \times \psi = 200 \times 0.5 = 50$, per capita recruitment $\gamma = 0.1$, annual survival $\phi = 0.8$, baseline capture probability $p_0 = 0.2$ and the scale parameter of the Gaussian capture probability function $\sigma = 50$. The initial $M=200$ activity centres were uniformly distributed within the square spatial region $S = [0, 1000] \times [0, 1000]$. We did not allow for movement of the activity centres over time.

The simulated populations were sampled at $J = 49$ traps, which were regularly spaced within a $[150, 850] \times [150, 850]$ square region. Hence, the minimum distance between a trap and the edge of S was 150 units, three times greater than σ . The number of secondary sampling occasions was $K = 5$, and the number of primary sampling occasions was $T = 10$.

We used seven subsets of each of the 100 data sets to evaluate the performance of our model. The seven scenarios are summarized in Table 3. The model was fitted to each of these 700 data sets, and posterior distributions were simulated by drawing 7000 MCMC samples, of which the initial 2000 were discarded. Sampling was conducted using JAGS, called from R, version 3.0.1 (Plummer 2003; R Core Team, 2013). We evaluated model performance by computing the root-mean-squared error and relative bias of the posterior mean estimates of annual abundance, and we also assessed coverage rate of 95% credible intervals of annual abundance.

SIMULATION RESULTS

The simulation study demonstrated that posterior precision was higher and bias was lower when combining SCR data

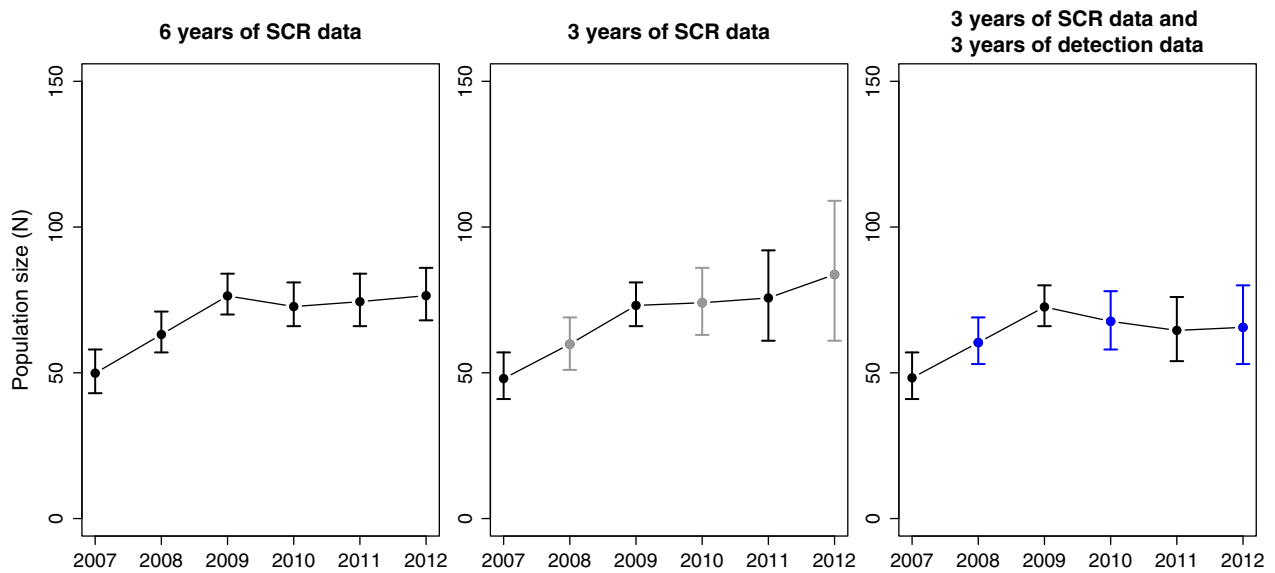


Fig. 3. Black bear population size estimates from the 3 scenarios. Black indicates years with SCR data, grey is years with no data, and blue is years with binary detection data.

Table 2. Posterior variances and means of N_t under the three bear analysis scenarios. **S1** makes use of all 6 years of SCR data, **S2** uses only SCR data from 2007, 2009 and 2011, and **S3** uses those 3 years of SCR data plus binary detection data in 2008, 2010 and 2012

Scenario	2007	2008	2009	2010	2011	2012
<i>Posterior variance</i>						
S1	15.53	13.07	12.23	14.11	19.76	22.31
S2	16.04	21.15	13.63	35.22	62.53	148.42
S3	15.64	16.69	12.37	24.23	33.43	48.25
<i>Posterior mean</i>						
S1	49.87	63.14	76.38	72.73	74.41	76.46
S2	48.02	59.83	73.12	74.03	75.68	83.70
S3	48.27	60.37	72.58	67.66	64.54	65.59

with simple detection data, relative to models based solely on the SCR data. For example, in scenario 7, the SCR data were only available in years 1, 5 and 9, but the use of the detection data in the other years resulted in 7.3–43% decreases in

RMSE relative to scenario 4 in which no data were available in years 2–4, 6–8 and 10 (Fig. 5). A similar pattern of decreased RMSE was also evident in the other scenarios, which are summarized in Appendix 3. Coverage of 95% credible intervals was nominal for all scenarios, ranging from 0.90 to 0.99 (Appendix 3).

We detected some positive bias in the estimators of annual abundance when applied to data sets that did not include SCR data in certain years (Appendix 3). The relative bias ranged from 0.17 to 2.1% for scenario 5, and from 3.0 to 4.6% for scenario 7. The bias was more pronounced for scenarios 2–4, which did not make use of the survey data. However, we found that this apparent bias resulted from convergence problems in a small number of data sets. These outliers are evident in Fig. 6, which depicts the worst case scenario for the data sets involving both SCR data and survey data. In spite of these outliers, the mode of the sampling distribution is closely aligned with the expected values of annual abundance indicating that the estimators are approximately unbiased.

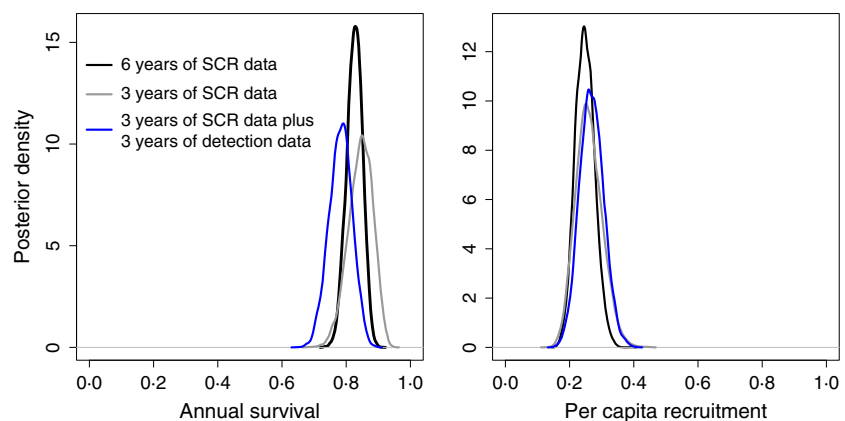
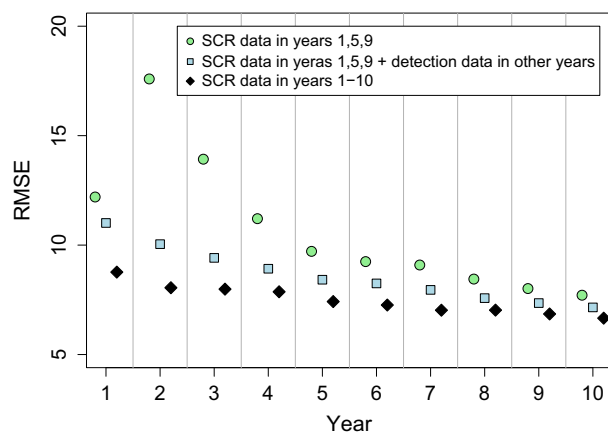


Fig. 4. Posterior distributions of annual survival (ϕ) and per capita recruitment (γ) of bears under the three scenarios described in the text and in Fig. 3.

Table 3. The seven simulation scenarios

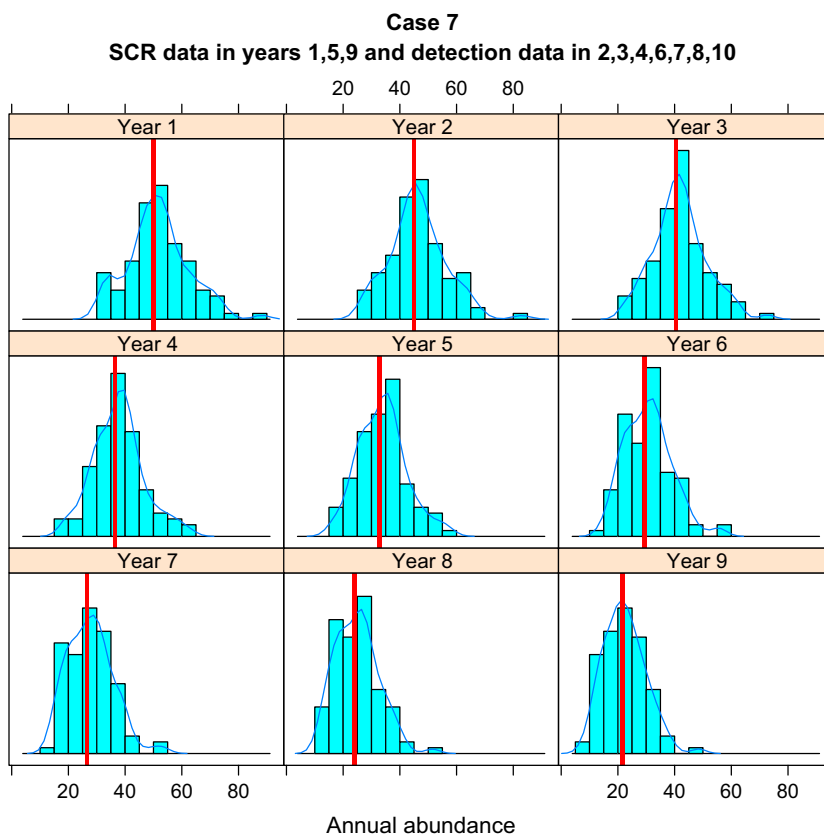
Scenario	Years with SCR data	Years with detection data
1	1–10	–
2	1,3,5,7,9	–
3	1,4,7,10	–
4	1,5,9	–
5	1,3,5,7,9	2,4,6,8,10
6	1,4,7,10	2,3,5,6,8,9
7	1,5,9	2,3,4,6,7,8,10

**Fig. 5.** Root-mean-squared error of the posterior means of annual abundance from simulation scenarios 1, 4 and 7. Models were fitted to 100 simulated data sets for each scenario.

Discussion

We developed a spatially explicit integrated population model that builds upon existing non-spatial models. Like previous IPMs, our model can be used to scale-up studies of the mechanisms driving population dynamics. This is accomplished by jointly analysing survey data, which can be collected at broad spatial and temporal scales, and capture–recapture data, which are more expensive to collect but contain more information about vital rates. As with non-spatial models, the spatial model allows for predictions of population size under future environmental conditions while accounting for process variation and observation error. However, unlike conventional IPMs, our model is parameterized entirely in terms of individuals and thereby allows for (spatially induced) individual heterogeneity in demographics and capture or detection probability. In addition, the model provides a coherent means of predicting abundance or density in unsampled regions and time periods, and thus can be used to map and project species distributions and vital rates. Furthermore, because the state model is a spatial point process, it avoids problems associated with scaling up continuously varying processes such as density or movement, which can introduce bias in non-spatial models (Wiens 1989, Melbourne & Chesson 2005).

Another appealing feature of our model is that the observation models for the capture–recapture data and the survey data are conditioned on the same underlying spatial model of population dynamics, avoiding the conventional requirement that the two data sets be independent. Mazzetta, Morgan & Coulson 2010 accomplished a similar objective using a single

**Fig. 6.** Histograms of annual abundance estimates (posterior means) obtained by fitting the model to 100 simulated data sets. The vertical lines indicate the expected values of abundance. Results from year 10 are omitted for compactness.

state model and two observation models for resighting data and mark–recapture–recovery data. From these two data sets, they were able to estimate seasonal and age-specific demographic parameters while accounting for imperfect sampling. Our model builds upon their work by allowing for spatial variation in each of the parameters.

One limitation of our model is that it is computationally demanding, especially when the number of individuals in the population is large. As an example, the MCMC sampling rate was slightly under 1000 iterations/hour with $M = 200$ for scenario S3 used in the black bear analysis. This problem can alleviate to some extent using parallel processing, and we expect computation time will decrease as computational power increases and novel analysis techniques are developed (e.g. Bonner & Schofield 2013).

In the analysis of the black bear data, we considered the situation in which capture–recapture data are only available in certain years, with survey data being collected in the other years. Our results indicate that the years with simple detection data do in fact provide information about population dynamics, even though, by themselves, these binary data could not be used to estimate demographic parameters. These results were demonstrated more conclusively by our simulation study. The increased posterior precision that we observed resulted from the fact that patterns in abundance and distribution are manifest in both the capture–recapture data and the detection data, a finding that agrees with previous theoretical and empirical work (He & Gaston 2003; Royle & Nichols 2003; Chandler & Royle 2013).

Some of the discrepancies we observed between the results from the three scenarios considered in the bear analysis were likely the result of model mis-specification since we intentionally ignored known sources of variation (e.g. sex, year and habitat effects) in demographic and detection parameters (Lowe 2011). Accommodating all of these sources of variation would have required a detailed model selection procedure that, while straight-forward, was beyond the scope of this paper. As a result, the estimates of abundance for this threatened species are likely to be biased high (although they are similar to previous estimates (O'Connell 2013)) and should not be used in management decisions.

The ability to restrict the expensive capture–recapture methods to a subset of years makes it possible to greatly increase the duration of a study for a fixed level of resources. In the bear example, the estimates of abundance obtained from S3 were similar to the estimates from S1 but laboratory expenses were halved, since the genotyping was not necessary. In addition to extending the temporal extent of a study, the model can also be used to increase the spatial extent. For example, it could be applied to designs in which the capture–recapture data are only collected in a subset of regions within the larger area of interest. This idea has been pursued by Conroy *et al.* (2008) in a static (i.e. no temporal variation in abundance) context. Like our model, their model includes an observation model describing how binary detections result from an underlying state model parameterized in terms of abundance. Thus, our model can be regarded as a spatially explicit extension of the Conroy *et al.*

(2008) model, and one that allows for variation in abundance over both time and space.

By allowing for population dynamics in space and time, both design options – collecting capture–recapture data in only a subset of regions and years – could be conducted simultaneously, raising the possibilities of highly efficient designs. Moreover, many existing data sets might be amenable to analysis with this model. For example, the North American Breeding Bird Survey (Sauer & Link, 2011) yields spatially-referenced count data and the Monitoring Avian Productivity and Survival (DeSante & Kaschube, 2009) project yields spatial capture–recapture data. [Correction added on 28 Feb 2014 after first online publication: line numbers removed that were included in error.] Our model could be applied to both data sets, thereby allowing for inference about abundance and demographics over large regions of North America. More generally, we believe this model represents an important step forward in efforts to scale-up inferences about the mechanisms governing population dynamics.

Acknowledgements

We thank the editor, associate editor, J. A. Royle, R. M. Dorazio, and two anonymous reviewers for suggestions that led to the improvement of the paper. We also wish to acknowledge C. Lowe and K. O'Connell for graciously allowing the use of data that they and many technicians collected in the field. Funding for this study was provided by the Louisiana Department of Wildlife and Fisheries, U.S. Fish and Wildlife Service, U.S. Geological Survey, U.S. Army Corps of Engineers, and Roy O. Martin Company.

References

- Abadi, F., Botha, A. & Altwegg, R. (2013) Revisiting the effect of capture heterogeneity on survival estimates in capture–mark–recapture studies: does it matter? *PLoS One*, **8**, e62636.
- Besbeas, P., Freeman, S.N., Morgan, B.J.T. & Catchpole, E.A. (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, B.J.T. (2003) The efficient integration of abundance and demographic data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **52**, 95–102.
- Best, N.G., Ickstadt, K. & Wolpert, R.L. (2000) Spatial Poisson regression for health and exposure data measured at disparate resolutions. *Journal of the American Statistical Association*, **95**, 1076–1088.
- Bonner, S. & Schofield, M. (2014) MC(MC)/MC: exploring Monte Carlo integration within MCMC for mark–recapture models with individual covariates. *Methods in Ecology and Evolution*, (In press).
- Borchers, D.L. & Efford, M.G. (2008) Spatially explicit maximum likelihood methods for capture–recapture studies. *Biometrics*, **64**, 377–385.
- Brooks, S.P., King, R. & Morgan, B.J.T. (2004) A Bayesian approach to combining animal abundance and demographic data. *Animal Biodiversity and Conservation*, **27**, 515–529.
- Buckland, S.T., Newman, K.B., Thomas, L. & Koesters, N.B. (2004) State-space models for the dynamics of wild animal populations. *Ecological Modelling*, **171**, 157–175.
- Chandler, R.B. & King, D.I. (2011) Habitat quality and habitat selection of golden-winged warblers in Costa Rica: an application of hierarchical models for open populations. *Journal of Applied Ecology*, **48**, 1038–1047.
- Chandler, R.B. & Royle, J.A. (2013) Spatially-explicit models for inference about density in unmarked or partially marked populations. *Annals of Applied Statistics*, **7**, 936–954.
- Clark, R.G., Hobson, K.A., Nichols, J.D. & Bearhop, S. (2004) Avian dispersal and demography: scaling up to the landscape and beyond. *The Condor*, **106**, 717–719.
- Clutton-Brock, T. & Sheldon, B.C. (2010) Individuals and populations: the role of long term, individual-based studies of animals in ecology and evolutionary biology. *Trends in Ecology & Evolution*, **25**, 562–573.

- Conroy, M.J., Runge, J.P., Barker, R.J., Schofield, M.R. & Fonnesbeck, C.J. (2008) Efficient estimation of abundance for patchily distributed populations via two-phase, adaptive sampling. *Ecology*, **89**, 3362–3370.
- de Valpine, P. & Hastings, A. (2002) Fitting population models incorporating process noise and observation error. *Ecological Monographs*, **72**, 57–76.
- DeSante, D.F. & Kaschube, D.R. (2009) The monitoring avian productivity and survivor ship (MAPS) program 2004, 2005, and 2006 report. *Bird Populations*, **9**, 86–169.
- Efford, M.G. (2004) Density estimation in live-trapping studies. *Oikos*, **106**, 598–610.
- Efford, M.G., Dawson, D.K. & Borchers, D.L. (2009) Population density estimated from locations of individuals on a passive detector array. *Ecology*, **90**, 2676–2682.
- Gardner, B., Reppucci, J., Lucherini, M. & Royle, J.A. (2010) Spatially explicit inference for open populations: estimating demographic parameters from camera-trap studies. *Ecology*, **91**, 3376–3383.
- Grosbois, V., Gimenez, O., Gaillard, J.-M., Pradel, R., Barbraud, C., Clobert, J., Møller, A. & Weimerskirch, H. (2008) Assessing the impact of climate variation on survival in vertebrate populations. *Biological Reviews*, **83**, 357–399.
- He, F. & Gaston, K.J. (2003) Occupancy, spatial variance, and the abundance of species. *The American Naturalist*, **162**, 366–375.
- Heidelberger, P. & Welch, P.D. (1983) Simulation run length control in the presence of an initial transient. *Operations Research*, **39**, 1109–1144.
- Lowe, C.L. (2011) Estimating population parameters of the Louisiana black bear in the Upper Atchafalaya River Basin. Master's thesis, University of Tennessee, Knoxville, TN.
- MacKenzie, D.I., Nichols, J.D., Royle, J.A., Pollock, K.H., Bailey, L.L. & Hines, J.E. (2006) *Occupancy Estimation and Modeling: Inferring Patterns and Dynamics of Species Occurrence*. Academic Press, Burlington, Massachusetts, USA.
- Mazerolle, D.F., Dufour, K.W., Hobson, K.A. & Haan, H.E.d. (2005) Effects of large scale climatic fluctuations on survival and production of young in a Neotropical migrant songbird, the yellow warbler *Dendroica petechia*. *Journal of Avian Biology*, **36**, 155–163.
- Mazzetta, C., Morgan, B.J.T. & Coulson, T. (2010) A state-space modelling approach to population size estimation. University of Kent Technical Report, UKC/SMSAS/10/025, 1–30.
- Melbourne, B.A. & Chesson, P. (2005) Scaling up population dynamics: integrating theory and data. *Oecologia*, **145**, 178–186.
- O'Connell, K. (2013) Population dynamics of the Louisiana black bear in the Upper Atchafalaya River Basin. Master's thesis, University of Tennessee, Knoxville, TN.
- Plummer, M. (2003) JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. March, pp. 20–22.
- Polansky, L., de Valpine, P., Lloyd-Smith, J.O. & Getz, W.M. (2009) Likelihood ridges and multimodality in population growth rate models. *Ecology*, **90**, 2313–2320.
- Pollock, K.H. (1982) A capture-recapture design robust to unequal probability of capture. *Journal of Wildlife Management*, **46**, 752–757.
- Post, E. & Forchhammer, M.C. (2002) Synchronization of animal population dynamics by large-scale climate. *Nature*, **420**, 168–171.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Royle, J.A. (2004) N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, **60**, 108–115.
- Royle, J.A., Chandler, R.B., Sollmann, R. & Gardner, B. (2013) *Spatial Capture-Recapture*. Academic Press, Waltham, Massachusetts, USA.
- Royle, J.A. & Dorazio, R.M. (2012) Parameter-expanded data augmentation for Bayesian analysis of pture models. *Journal of Ornithology*, **152**, 521–537.
- Royle, J.A., Dorazio, R.M. & Link, W.A. (2007) Analysis of multinomial models with unknown index using data augmentation. *Journal of Computational and Graphical Statistics*, **16**, 67–85.
- Royle, J.A. & Nichols, J.D. (2003) Estimating abundance from repeated presence-absence data or point counts. *Ecology*, **84**, 777–790.
- Sauer, J.R. & Link, W.A. (2011) Analysis of the North American breeding bird survey using hierarchical models. *The Auk*, **128**, 87–98.
- Schaub, M. & Abadi, F. (2011) Integrated population models: a novel analysis framework for deeper insights into population dynamics. *Journal of Ornithology*, **152**, 227–237.
- Schaub, M., Gimenez, O., Siero, A. & Arlettaz, R. (2007) Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. *Conservation Biology*, **21**, 945–955.
- Schaub, M. & Royle, J.A. (2014). Estimating true instead of apparent survival using spatial Cormack-Jolly-Seber models. *Methods in Ecology and Evolution*. In press.
- Sollmann, R., Gardner, B., Parsons, A.W., Stocking, J.J., McClintock, B.T., Simons, T.R., Pollock, K.H. & O'Connell, A. (2012) A spatial mark-resight model augmented with telemetry data. *Ecology*, **94**, 553–559.
- U.S. Fish and Wildlife Service (1992). Determination for threatened status for *U. a. luteolus* (Louisiana black bear). *Federal Register*, **57**, 588–595.
- U.S. Fish and Wildlife Service (1995) *Louisiana Black Bear Recovery Plan*. Jackson, Mississippi, USA.
- Wiens, J. A. (1989) Spatial scaling in ecology. *Functional Ecology*, **3**, 385–397.
- Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and Management of Animal Populations: Modeling, Estimation, and Decision Making*. Academic Press.

Received 19 September 2013; accepted 9 December 2013

Handling Editor: Evan Cooch

Supporting Information

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

Appendix S1. Field and Lab Methods.

Appendix S2. JAGS code to fit the model to the bear data.

Appendix S3. Results from simulation study involving 100 datasets and 7 scenarios.