Estimating true instead of apparent survival using spatial Cormack–Jolly–Seber models

Michael Schaub[1]* and **J. Andrew Royle[2]**

[1]*Swiss Ornithological Institute, CH–6204 Sempach, Switzerland; and* [2]*USGS Patuxent Wildlife Research Center, 12100 Beach Forest Road, Laurel, ML 20708, USA*

Summary

1. Survival is often estimated from capture–recapture data using Cormack–Jolly–Seber (CJS) models, where mortality and emigration cannot be distinguished, and the estimated apparent survival probability is the product of the probabilities of true survival and of study area fidelity. Consequently, apparent survival is lower than true survival unless study area fidelity equals one. Underestimation of true survival from capture–recapture data is a main limitation of the method.

2. We develop a spatial version of the CJS model that allows estimation of true survival. Besides the information about *whether* a specific individual was encountered at a given occasion, it is often recorded *where* the encounter occurred. Thus, information is available about the fraction of dispersal that occurs within the study area, and we use it to model dispersal and estimate true survival. Our model is formulated hierarchically and consists of survival, dispersal and observation submodels, assuming that encounters are possible anywhere within a study area.

3. In a simulation study, our new spatial CJS model produced accurate estimates of true survival and dispersal behaviour for various sizes and shapes of the study area, even if emigration is substantial. However, when the information about dispersal is scarce due to low survival, low recapture probabilities and high emigration, the estimators are positively biased. Moreover, survival estimates are sensitive to the assumed dispersal kernel.

4. We applied the spatial CJS model to a data set of adult red-backed shrikes (*Lanius collurio*). Apparent survival of males (*c*. 0·5) estimated with the CJS model was larger than in females (*c*. 0·4), but the application of the spatial CJS model revealed that both sexes had similar survival probabilities (*c*. 0·6). The mean breeding dispersal distance in females was *c*. 700 m, while males dispersed only *c*. 250 m between years.

5. Spatial CJS models enable study of dispersal and survival independent of study design constraints such as imperfect detection and size of the study area provided that some of the dispersing individuals remain in the study area. We discuss possible extensions of our model: alternative dispersal models and the inclusion of covariates and of a habitat suitability map.

Key-words: Cormack–Jolly–Seber model, dispersal, *Lanius collurio*, movement, sampling bias, spatial capture–recapture

Introduction

Survival is a key demographic process for the dynamics of populations. Knowledge about the magnitude of survival, how it varies temporally and spatially and by which factors temporal and spatial variation is induced are essential topics in demographic studies. Survival is very often estimated from capture–recapture data: individuals are marked and released into the study population and are then followed through time by re-encountering them. Since detection is not perfect, the resulting capture–recapture data are typically analysed with the Cormack–Jolly–Seber (CJS) model (see Lebreton *et al.* 1992; Williams, Nichols & Conroy 2002), which allows the separate estimation of the probabilities of survival and recapture. How-

ever, if individuals emigrate from the study population, these parameters may be biased and the bias depends on the type of emigration. If emigration is temporary and random, only recapture probabilities are biased, while temporary non-random emigration will bias both parameters (Schaub *et al.* 2004). Finally, if emigration is permanent, the estimated survival probability is biased low as mortality and emigration cannot be distinguished with CJS models. The estimated parameter is referred to as the apparent survival probability which is the product of the probabilities of true survival and of study area fidelity (Lebreton *et al.* 1992). Consequently, apparent survival is lower than true survival unless study area fidelity equals one. The underestimation of true survival from capture–recapture data is a major limitation of the method.

The degree of underestimation of true survival by apparent survival can be substantial (Cilimburg *et al.* 2002; Marshall

*Correspondence author. E-mail: michael.schaub@vogelwarte.ch

et al. 2004) and depends on the size of the study area and on the dispersal behaviour of the species under study (Zimmerman, Gutierrez & LaHaye 2007). If the study area is large relative to dispersal, a majority of dispersing individuals will settle within the study area and apparent survival approximates true survival well. Thus, the degree of underestimation depends on a combination of the study design (specifically, study area) and the dispersal behaviour. This is worrying for several reasons. First, different groups or age classes of individuals in the study population may have different dispersal behaviour resulting in different estimates of apparent survival, even if true survival does not vary by group or age class. Therefore, it is risky to test whether true survival differs among groups or age classes. Secondly, if the temporal variation of survival is of interest, it is unclear how much the observed pattern reflects temporal variation in true survival, how much of it reflects variation in site fidelity or of a combination of both. Thirdly, if spatial variation of survival is studied by comparing apparent survival from different study areas, the estimated variation could stem from variation in true survival or from variation in site fidelity or a combination of both, which might be due to differences in the study area size and shape. For these reasons, it would be desirable to estimate true instead of apparent survival also from capture–recapture data.

Some attempts have been made to reduce the negative bias of survival and all of them need additional data. The most promising ones are those where the capture–recapture data are analysed jointly with data that allow the estimation of true survival, such as band-recovery data (Burnham 1993), observations of marked individuals outside the study area (Barker 1997) or telemetry data (Powell *et al*. 2000). If dispersal takes place among just a few well-defined sites (e.g. colonies in seabirds), sampling at all of these sites and the use of multistate capture–recapture models allow the estimation of true survival (Lebreton *et al*. 2009). Captures/resightings outside the core study area can also help to reduce the bias (Marshall *et al*. 2004). Gilroy *et al*. (2012) developed recently an interesting model that explicitly used information about the locations of encounters within the study area. Using a two-step approach, they generated a spatial projection of dispersal probability around each capture location which is used in a second step to estimate permanent emigration probability and finally survival. Finally, Ergon & Garnder (2014) developed a spatial robust design model that can be applied to a sampling design with fixed trapping locations. The model uses the spatial information about the trapping locations and allows the estimation of dispersal and true survival.

Here, we develop a spatially explicit capture–recapture model for open populations that allow the estimation of true survival, that is, a spatial generalization of the CJS model. Besides the information about *whether* a specific individual was encountered at a given occasion, it is often recorded *where* the encounter occurred. A fraction of dispersal, namely dispersal that occurs within the study area, can be observed. This information can be used to gauge emigration from the study area and thus estimate true survival. Our model is based on a similar idea as is the model by Gilroy *et al*. (2012). However, in contrast

to their model, we estimate and model dispersal and survival jointly within a single unified hierarchical model. This has the advantages that estimation errors are fully accounted for and of much increased flexibility in modelling. Our model shows similarities with the model of Ergon & Garnder (2014). A main difference is that they considered a different sampling design with fixed trapping locations while we assumed that recapture is possible anywhere within a defined study area. We introduce our model, present a simulation study to assess its performance and finally apply it to a real data set on a passerine bird.

Materials and methods

We require capture–recapture data from a study area and summarize them in a matrix **Y**. The element $Y_{i,t}$ of **Y** equals 1 if individual *i* is captured at time *t*, and 0 otherwise. In addition, the locations (coordinates) of each capture are needed, and we store this information in an array **G**. For individual *i* captured at time *t*, $G_{i,t,1}$ is the *x*-coordinate and $G_{i,t,2}$ is the *y*-coordinate of the location of capture. When individual *i* is not captured at *t*, its location is obviously unknown, and thus, $G_{i,t,1}$ and $G_{i,t,2}$ are missing values. We assume that the locations are recorded without measurement errors, that encounters are possible anywhere within the study area and that the study area is defined as the area in which animals are marked and re-encountered.

The model that we propose is a generalized CJS model, where the extension is represented by an additional submodel for the dispersal process. Hence, we call our model a spatial CJS model. To implement the model, we use a state-space formulation (Gimenez *et al*. 2007; Royle 2008; Kéry & Schaub 2012). Our model consists of two state processes that evolve in a Markovian fashion and one observation process. Survival is modelled in the first state process. We define a latent state variable $z_{i,t}$ to indicate whether individual *i* is alive ($z_{i,t} = 1$) or dead ($z_{i,t} = 0$) at time *t*. Note that this definition is different from the definition of the state variable of the non-spatial CJS model, where $z_{i,t}$ is an indication of whether individual *i* is alive at time *t and* has not permanently left the study area (Gimenez *et al*. 2007; Royle 2008; Kéry & Schaub 2012). The CJS model conditions on first capture; hence, we denote with f_i the occasion at which an individual is marked. Then,

$$z_{i,f_i} = 1$$
$$z_{i,t+1} | z_{i,t} \sim \text{Bernoulli}(z_{i,t} s_{i,t}) \qquad \text{eqn 1}$$

where $s_{i,t}$ is the true survival probability of individual *i* between *t* and *t* + 1. We make the assumption that survival does not vary spatially, that is, we assume that survival is identical inside and outside the study area. We further assume that $z_{i,t}$ are independent among individuals, conditional on survival probabilities, $s_{i,t}$, so that the joint distribution of all $z_{i,t}$ variables is the product

$$\prod_i \left(\prod_{t=f_i+1} [z_{i,t} | z_{i,t-1}, s_{i,t}] \right) z_{i,f_i}.$$

The second state process model describes the dispersal location variables **G**. Depending on the dispersal behaviour of an

individual, **G** may change from occasion to occasion. We describe dispersal as a random walk (Turchin & Thoeny 1993; Ovaskainen 2004) and model the location of individual i at time $t + 1$ with a normal distribution, whose mean is the location of individual i at t and a variance that is estimated:

$$\mathbf{G}_{i,t+1} \sim \text{Normal}\left(\mathbf{G}_{i,t}, \sigma_G^2 \mathbf{I}\right), \qquad \text{eqn 2}$$

where σ_G^2 are dispersal variances in x- and in y-direction (i.e. σ_G^2 is a vector with two values). Thus, we do not model directly dispersal distance, but instead we estimate dispersal variances in two directions. The expected dispersal distance is related to the dispersal variances as $D = \sqrt{2\left(\sigma_x^2 + \sigma_y^2\right)/\pi}$.

The final step of the model formulation is the observation process. Only individuals that are alive and present in the study area are available for capture, that is, the observation process is conditional on survival and presence in the study area. We denote with **A** the state-space of the study area and with $r_{i,t}$ an indicator of whether individual i is inside or outside of the study area at time t:

$$r_{i,t} = 1, \text{ if } \mathbf{G}_{i,t} \in \mathbf{A}$$

$$r_{i,t} = 0, \text{ if } \mathbf{G}_{i,t} \notin \mathbf{A}$$

The observation model is then:

$$y_{i,t} | z_{i,t}, r_{i,t} \sim \text{Bernoulli}\left(z_{i,t} r_{i,t} p_{i,t}\right) \qquad \text{eqn 3}$$

where $p_{i,t}$ is the recapture probability of individual i at time t.

The defined model is parameter redundant and needs some constraints (e.g. the same in all individuals or constant over time) such that the parameters are separately identifiable. We assume here that survival and recapture probabilities are the same in all individuals and that they are constant over time. Because the model uses a normal distribution for modelling the dispersal process, we call it a spatial CJS model with normal dispersal and denote it by sCJS-N.

SIMULATION STUDY

We evaluated the performance of the sCJS-N model with a simulation study. In a first setting, we evaluated the performance of the model for different values of dispersal, survival and recapture each for a square study area with 15 km side length and for an irregularly shaped study area consisting of 25 square grid cells with 2 km side length each (see Appendix S5, Fig. S5.1, Supporting Information shape 1). We considered 16 scenarios that differ in mean survival, recapture and dispersal variance. We varied the dispersal variance ($\sigma_G^2 = \{0.5, 5, 15, 50\}$ in x and y directions) corresponding to mean dispersal distances of about 0·88, 2·80, 4·85 and 8·89 km, respectively. We varied the recapture probability using $p = 0.4$ and $p = 0.8$, and true survival $s = 0.4$ and $s = 0.7$. We assumed that all parameters are constant over time. In all scenarios, we used seven study years and assumed that 100 newly marked individuals are released at each occasion but the last one, resulting in a total number of 600 capture histories.

In a second setting, we assessed whether the sCJS-N model is able to produce accurate estimates irrespective of the size and the shape of the study area. To study the impact of the size of the study area, we considered square study areas with side lengths 10, 15 and 20 km, corresponding to sizes of 100, 225 and 400 km². To study the impact of the shape of the study area, we designed three different study areas consisting of 25 square grid cells with 2 km side length each, but the grid cells were spatially arranged different resulting in differentially shaped study areas of the same size (100 km², see Appendix S5, Fig. S5.1, Supporting Information). In each of these settings, we varied the dispersal variance as before ($\sigma_G^2 = \{0.5, 5, 15, 50\}$ in x and y directions), but we always used recapture probability of $p = 0.8$ and true survival of $s = 0.7$. Again we used parameters that were constant over time, seven study years and released 100 newly marked individuals at each but the last occasion.

To simulate the data, we first randomly generated the location of each newly marked individual in the study area. For subsequent occasions, we computed whether an individual has survived using a Bernoulli distribution with success probability s. For survivors, the new location was randomly drawn from Normal distributions as in eq. 2. Next, we determined whether an individual was captured using another Bernoulli trial with success probability p. This was repeated until an individual had either died or the study ended. We assembled the capture–recapture data of the individuals recorded within the study area in a matrix **Y** and the observed positions in array **G**. We also recorded the number of recaptures and the proportion of individuals alive that are presented outside the study area. This proportion is an estimate of the emigration probability.

We conducted two Bayesian analyses with these data sets: first, we analysed **Y** and **G** with the sCJS-N model, and secondly, we analysed **Y** with a conventional CJS model. We used vague priors for all parameters ($s \sim U(0,1)$; $p \sim U(0,1)$; $\sigma_G \sim U(0,50)$). We conducted 200 simulation replicates for each scenario and report absolute and relative bias, mean squared error (MSE) and coverage rate (the proportion of times the true value was included in the central 95% credible interval) of the posterior means of s, p and σ_G^2. Note that the evaluation of the estimators is a frequentist one. We have chosen this approach since our interest was to understand the performance of the estimators in specific regions of the parameter space, that is, conditional on specific values of parameters that we think are typical in bird populations.

Data simulation was performed in R 2.14.0 (R Development Core Team 2004), and the data were analysed with JAGS (Plummer 2003) executed from R via the package R2jags. For each model, we ran two Markov chains initiated at random starts for 10 000 iterations. We discarded the first 5000 samples as burn-in and kept the remainder to summarize the posterior distributions. We only used simulations with chains that had converged, that is, had an $\hat{R} < 1 \cdot 1$ for all parameters (Brooks & Gelman 1998). We therefore performed as many simulations as were necessary to accumulate 200 valid estimates of each parameter and each scenario under each model. Simulation and analysis code is available in Appendices S1 and S2 (Supporting Information).

CASE STUDY: SURVIVAL OF RED-BACKED SHRIKES

We illustrate our spatial CJS model for the estimation of true survival and dispersal distances using a data set of adult red-backed shrikes (*Lanius collurio*). This migratory passerine bird species is widespread in Europe and breeds in bushes or hedgerows on farmland. The capture–recapture data were collected from 1988 to 1992 always during the breeding period lasting from mid-May to end of July in a $2·45$ km^2 Swiss study area with irregular shape (Ramosch, 46°50′N/10°23′E, 1090–1680 m a.s.l.). Unmarked shrikes were captured with clap or mist nests and colour-marked. The study area was visited several times each year to resight colour-marked surviving shrikes and to locate their nests. The locations of the nests were taken as the spatial reference of the sightings. A total of 112 adult males and 94 adult females were marked of which 53 males and 31 females were resighted in at least 1 year. See Pasinelli *et al.* (2007) for a detailed description of study area and the sampling protocol.

Breeding dispersal of red-backed shrikes is not unusual and depends in the first place on the breeding success of the previous year (Pasinelli *et al.* 2007; Schaub, Jakober & Stauber 2011). Since apparent survival typically differs between sexes in red-backed shrikes due to greater dispersal of females (Schaub, Jakober & Stauber 2011), we modelled survival and dispersal variance separately for each sex. Due to the short duration of the study, we assumed that all parameters in the model were constant over time. We fitted three models to the data. First, we used the model with normal dispersal model (sCJS-N). Since the study area has an irregular shape, we subdivided it into 507 square grid cells with 70 m side length. The assignment test of whether the location of individual i at time t ($\mathbf{G}_{i,t}$) was inside the study area was performed for each grid cell. If $\mathbf{G}_{i,t}$ was located within one of the grid cells, $r_{i,t} = 1$, if not, $r_{i,t} = 0$. Secondly, we used a model that uses a *t*-distribution for modelling movements in both the *X* and *Y* directions, allowing more heavily tailed dispersal distance distributions (sCJS-T). This model is $\mathbf{G}_{i,t+1} \sim t\left(\mathbf{G}_{i,t}, \sigma_G^2 \mathbf{I}, \mathbf{d.f.}\right)$ which has additional parameters, the degrees of freedom (d.f.) for *X* and *Y*. We used a distribution that supports larger values than 2 as prior for both degrees of freedom. This prior was implemented indirectly by specifying $k \sim U(0, 0·5)$ and calculating d.f. $= 1/k$. The same assignment test as for the sCJS-N model was used. Finally, we fitted the ordinary CJS model for comparison.

We used the deviance information criterion (DIC) to evaluate whether the sCJS-N or the sCJS-T model was better supported by the data. We report posterior means and central 95% credible intervals (CRI) of the estimates. We then computed the proportion of marked individuals that settled outside the study area and the dispersal kernels (frequency distribution of dispersal distances) from the estimated locations ($\mathbf{G}$).

Results

SIMULATION STUDY

The simulations of the 16 scenarios for the square study area created a wide range of numbers of recaptured individuals and of emigration probabilities (Appendix S3, Table S3.4, Supporting Information). The mean number of recaptures ('sample size') decreased from 663 to 49 with decreasing survival and recapture probabilities and with increasing dispersal variance. The mean proportion of individuals alive outside the study area increased from 0·10 to 0·70 with increasing dispersal variance and with increasing survival. Sample size increased with increasing size of the study area while emigration decreased (Appendix S5, Table S5.4, Supporting Information).

Survival probability estimated with the sCJS-N model was virtually unbiased (maximal relative bias: 0·7%) in the scenarios with high survival ($s = 0·7$) irrespective of the chosen dispersal variances and recapture probabilities (Fig. 1, Appendix S3, Table S3.1, Supporting Information). In the scenarios with low survival ($s = 0·4$), survival was hardly biased (maximal relative bias: 1·5%) for dispersal variance ≤ 15 irrespective of the chosen levels of recapture. However, when dispersal variance was 50, the estimated survival tended to be slightly positively biased (maximal relative bias: 6%). The MSEs of the survival probabilities were low, but increased with increasing dispersal variances and decreasing survival and recapture probabilities (Appendix S3, Table S3.1, Supporting Information). The coverage rate was close to 0·95 in all scenarios (Appendix S3, Table S3.1, Supporting Information). The CJS model always underestimated survival, as expected (Fig. 1). The negative bias increased with increasing dispersal variance and became substantial (relative bias of up to 30%). The MSE of survival increased with increasing dispersal variance and with decreasing survival and recapture probabilities. It was typically much larger than the MSE of survival from the sCJS-N model (Appendix S3, Table S3.1, Supporting Information). Coverage rates were <0·95 in all scenarios and decreased with increasing dispersal variance (Appendix S3, Table S3.1, Supporting Information).

The estimated dispersal variances showed only little bias for dispersal variance ≤ 15 (maximal relative bias 5·5%, Fig. 2, Appendix S3, Table S3.2, Supporting Information). When the dispersal variance was 50, the relative bias became stronger (up to 40%) and it was always positive. The bias declined with increasing recapture and survival probabilities. The MSE of the dispersal variance increased with increasing dispersal variance. Unless the dispersal variance was very large, the coverage rate was close to 0·95 (Appendix S3, Table S3.2, Supporting Information).

The recapture probabilities estimated by fitting the sCJS-N model showed little bias for dispersal variance ≤ 15 (Fig. 3, Appendix S3, Table S3.3, Supporting Information). Recapture probabilities were strongly positively biased for dispersal variance of 50. The bias was particularly large when recapture
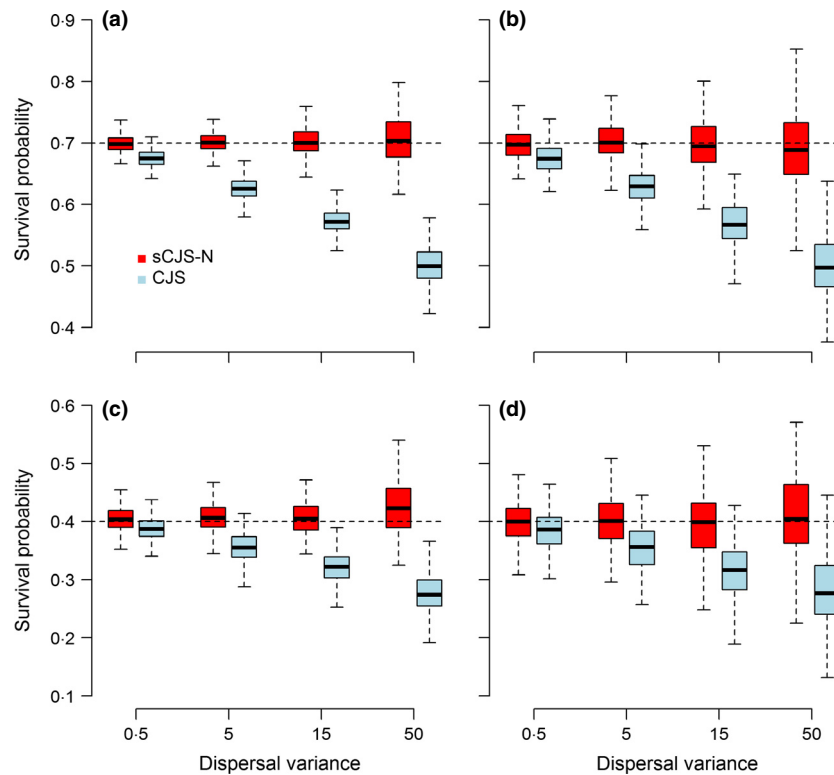
Fig. 1. Boxplots of the posterior means of survival probability ($n = 200$ simulations) in relation to dispersal variance in a square study area with 15 km side length. Recapture probabilities were 0·7 (panels a and c) and 0·4 (panels b and d). The dotted lines show the level of survival used in the simulation set-ups. Boxplots show the median (fat line in the centre of the box), the 25% and 75% quantiles (limits of boxes) and ranges within 1·5 times the height of the box (whiskers).
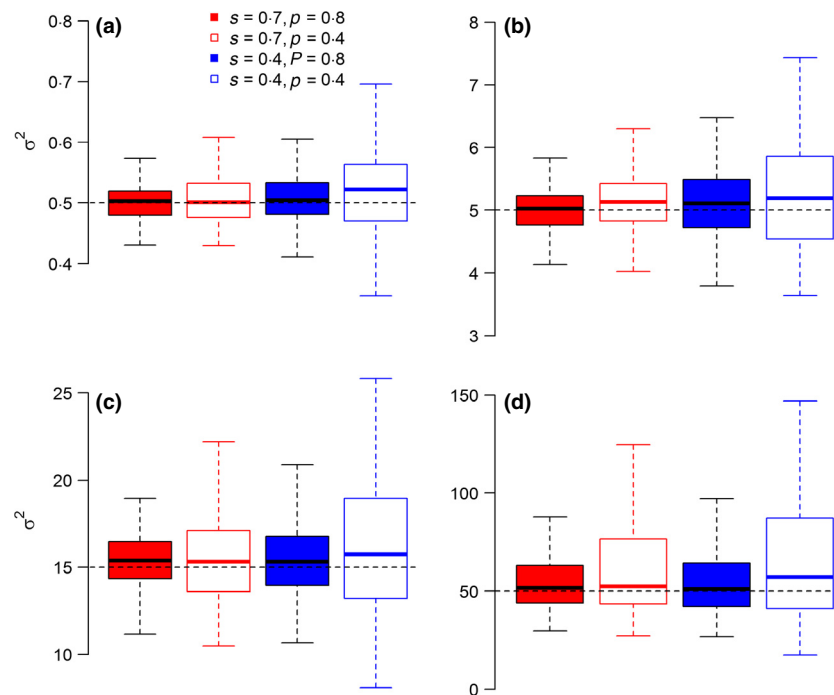

Fig. 2. Boxplots of the posterior means of dispersal variance ($n = 200$ simulations) in relation to different levels of survival and recapture probabilities in a square study area with 15 km side length. The four panels refer to different levels of dispersal variances used in the simulations. The dotted lines show the level of dispersal variances used in the simulation set-ups. Boxplots show the median (fat line in the centre of the box), the 25% and 75% quantiles (limits of boxes) and ranges within 1·5 times the height of the box (whiskers).
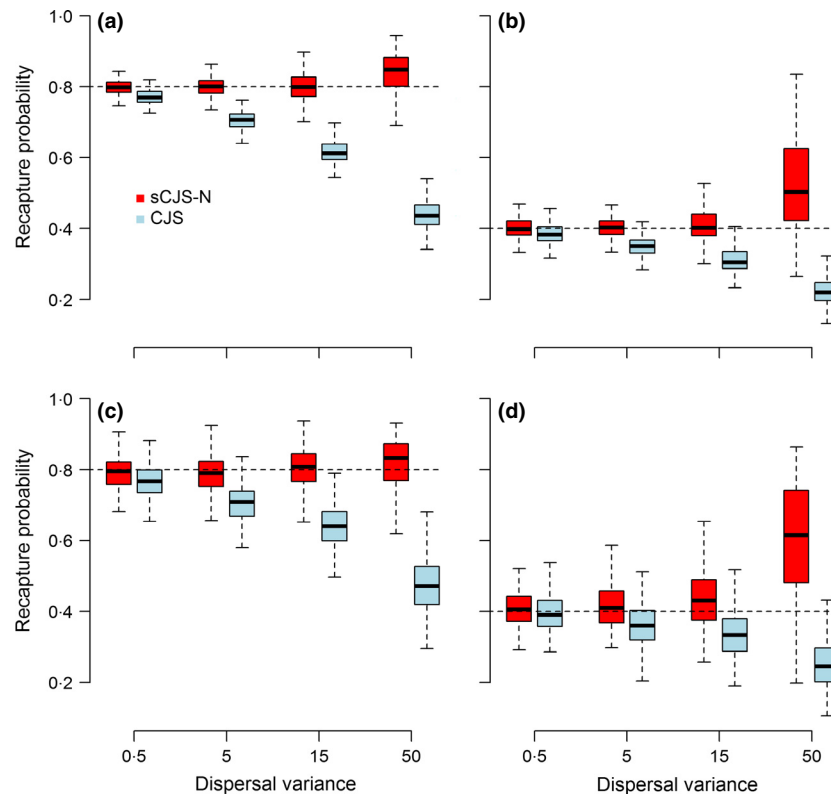
Fig. 3. Boxplots of the posterior means of recapture probability (*n* = 200 simulations) in relation to dispersal variance in a square study area with 15 km side length. Survival probabilities were 0·7 (panels a and b) and 0·4 (panels c and d). The dotted lines show the level of recapture probabilities used in the simulation set-ups. Boxplots show the median (fat line in the centre of the box), the 25% and 75% quantiles (limits of boxes) and ranges within 1·5 times the height of the box (whiskers).

probability was low (*p* = 0·4). The coverage rate was close to 0·95 in most scenarios, but when recapture probability was low and the dispersal variance large, it was <0·95. The relative bias of the recapture probability estimated with the CJS model was substantial and always negative. This was expected as temporary emigration results in negatively biased estimates of recapture probability in the traditional CJS model (Schaub *et al.* 2004). The coverage rate was often much smaller than 0·95 (Appendix S3, Table S3.3, Supporting Information).

Similar results were obtained when the study area was irregularly shaped (Appendix S4, Supporting Information). Most importantly, survival was virtually unbiased for different levels of dispersal variances. Generally, the MSEs of the estimates were larger compared to the MSE of the estimates from the square study area. This can be explained by the smaller number of recaptured individuals in the irregularly shaped study area due to a smaller size of the study area.

Survival probabilities estimated with the sCJS-N model were virtually unbiased for different sizes and shapes of study areas (Fig. 4). The MSE of survival increased with increasing dispersal variance and with decreasing size of the study areas (Appendix S5, Table S5.1, Supporting Information). The negative bias of survival estimated with the CJS model became stronger, the smaller the study area was (Fig. 4a–c), while different shapes of equally sized study areas did not result in strikingly different bias (Fig. 4d–f). Estimates of the dispersal variances and of recapture proba-

bilities under the sCJS-N model were also fairly unbiased and accurate (Appendix S5, Figs S5.2 and S5.3, Tables S5.2 and S5.3, Supporting Information). Recapture probabilities estimated with the CJS model were negatively biased and the bias increased with increasing dispersal variance and declining size of the study area (Appendix S5, Fig. S5.3, Table S5.3, Supporting Information). In summary, these simulations show that the sCJS-N model produces accurate parameter estimates regardless of the size and shape of the study area (i.e. a design aspect of the study).

CASE STUDY: SURVIVAL OF RED-BACKED SHRIKES

Capture locations and observed dispersal events are shown in Fig. 5. Estimates of the sex-specific survival probabilities were highest under the sCJS-N model and, as expected, lowest under the CJS model (Fig. 6). Survival probabilities of the two sexes were very similar under the sCJS-N (probability that survival of males was higher than that of females: 0·47) and the sCJS-T models (probability of 0·59), but male survival probability was higher than female survival under the CJS model (probability of 0·98). The DIC was much larger for the sCJS-N model (613) than for the sCJS-T model (468), indicating that the latter was better supported by the data. The resighting probabilities estimated with the sCJS-T model were high and similar for both sexes (males: 0·90 [CRI: 0·76–0·99]; females: 0·92 [CRI: 0·74–0·99]).
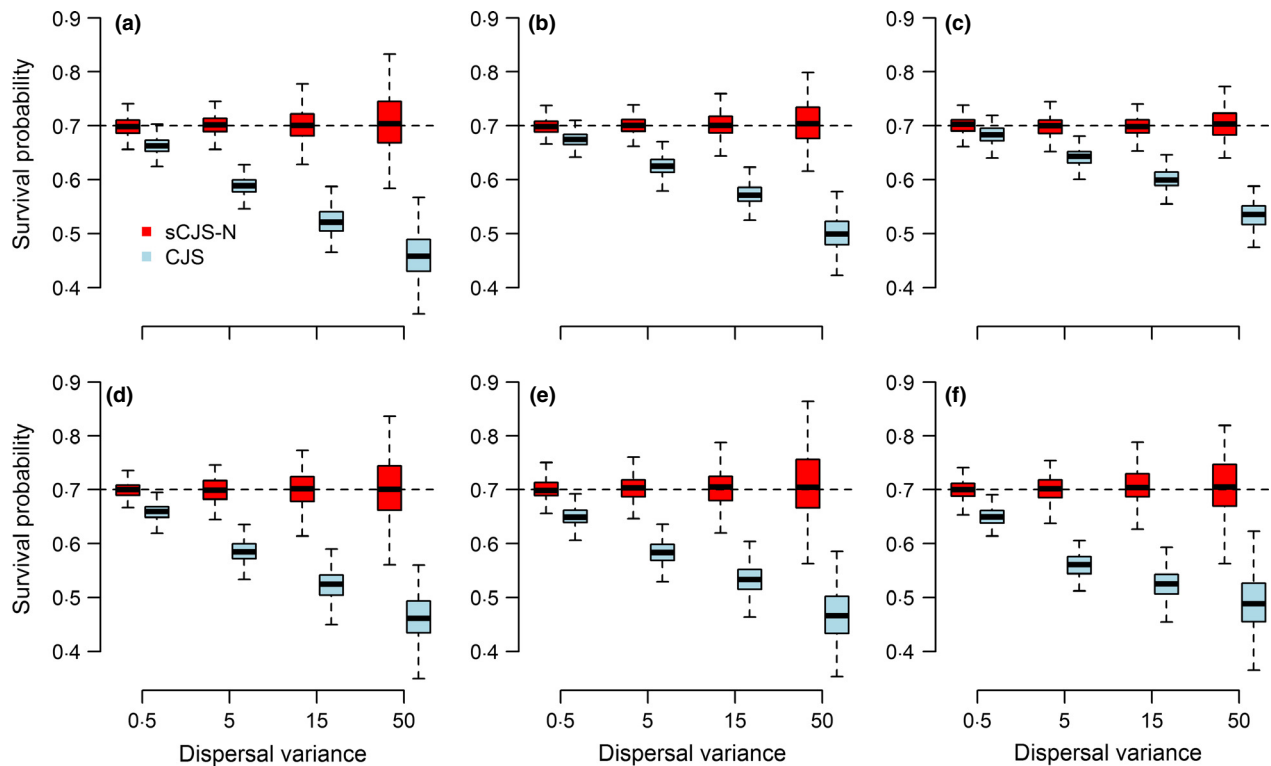
Fig. 4. Boxplots of the posterior means of survival probability (*n* = 200 simulations) in relation to dispersal variance in study areas of different sizes and shapes. Panels a, b and c show the results from square study areas with side lengths 10, 15 and 20 km, respectively. Panels d, e and f show the results from study areas of 100 km^2, but each of them had a different shape (see Appendix S5, Fig. S5.1, Supporting Information for the shapes). Survival probability was 0·7 and recapture probability 0·8 in all panels. Boxplots show the median (fat line in the centre of the box), the 25% and 75% quantiles (limits of boxes) and ranges within 1·5 times the height of the box (whiskers).
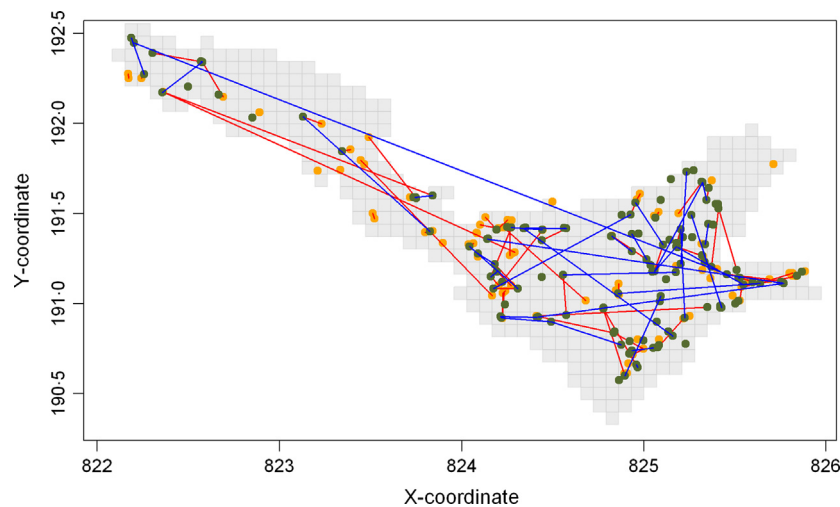

Fig. 5. Red-backed shrike study area in Switzerland and the location of marked and resighted individuals. Orange dots refer to males, green dots are females. Observed dispersal events are connected with lines (males: red; females: blue).

We used the sCJS-T model to make inference on dispersal. The proportion of shrikes that settled outside the study area in the years after marking was on average 0·27 (CRI: 0·09–0·45) in males and 0·50 (CRI: 0·25–0·69) in females. The estimated dispersal kernel for each sex is shown in Fig. 7. Mean dispersal distances of males were 0·25 km (CRI: 0·01–1·10 km); only about 10% of the dispersal distances were longer than 0·5 km. Females dispersed over longer distances (mean: 0·68 km; CRI: 0·01–

2·61 km) and about half of the dispersal distances were >0·5 km. The observed dispersal distances (males: 0·20 km [CRI: 0·01–1·15], *n* = 73; females: 0·44 [CRI: 0·01–1·74], *n* = 39) were shorter than the estimated ones in both sexes. Observed dispersal distances longer than 0·5 km were infrequent in both sexes (males: 7%; females: 28%). In conclusion, our spatial CJS model provided evidence that annual survival of adult red-backed shrikes is similar in both sexes and females dispersed over
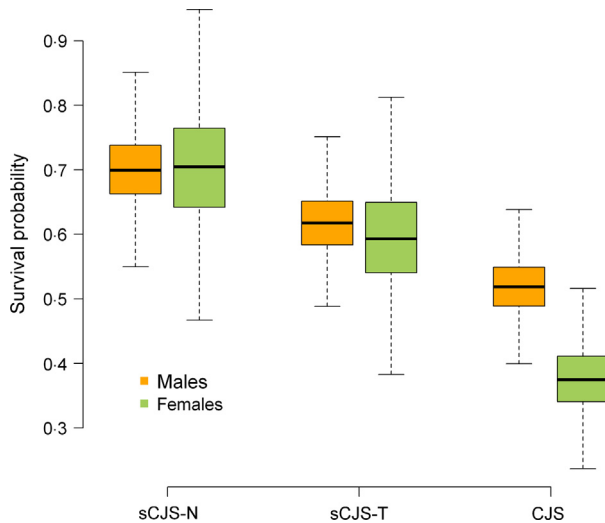
Fig. 6. Boxplot of the posterior distributions of the sex-specific survival probabilities of red-backed shrikes estimated under three different models (sCJS-N: spatial Cormack–Jolly–Seber (CJS) model with normally distributed dispersal, sCJS-T: spatial CJS model with *t*-distributed dispersal, CJS: ordinary CJS model). Boxplots show the median (fat line in the centre of the box), the 25% and 75% quantiles (limits of boxes) and ranges within 1·5 times the height of the box (whiskers).
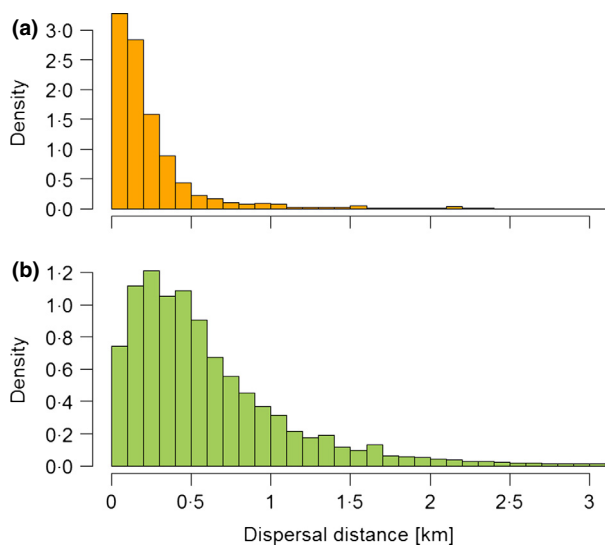

Fig. 7. Frequency distribution of dispersal distances (dispersal kernels) of red-backed shrike males (panel a) and female (panel b) calculated from the estimated locations under model sCJS-T.

greater distances than males. Differential dispersal resulted in apparent survival probabilities that differed between sexes.

Discussion

We have developed a spatially explicit version of the well-known CJS model (Lebreton *et al.* 1992) that allows the estimation of true survival regardless of dispersal or the size and the shape of a study area. Our hierarchical model extends the CJS model by the inclusion of a dispersal model. The CJS model can be seen as a special case of the spatial capture–recapture model developed here, when dispersal is assumed to

be absent. Our simulation study suggests that the new model performs well for a wide range of dispersal movements and for different sizes and shapes of study areas given that the model assumptions are fulfilled. Survival is estimated without or with only little positive bias, whereas the CJS model, as well known, seriously underestimates survival when dispersal movements beyond the study area occur (Cilimburg *et al.* 2002; Marshall *et al.* 2004). The application of the new model to a data set of red-backed shrikes has shown that their survival is very similar in both sexes and that dispersal is stronger in females than in males. This result is in accordance with current knowledge about dispersal and survival of red-backed shrikes (Pasinelli *et al.* 2007; Schaub, Jakober & Stauber 2013). It contrasts sharply with the traditional CJS estimates which suggest lower apparent survival in females compared to males. Thus, we obtained improved ecological insights by the application of the spatially explicit CJS model.

A main advantage of our model is that estimates of both true survival and of dispersal are obtained. Both demographic processes are interesting on their own and are important components for an understanding of how a population works. The model offers a framework for modelling effects of time, individual or age on survival and/or dispersal via linear modelling (Kéry & Schaub 2012). Typically, study areas are defined based on logistic constraints rather than on biological grounds (Thomas & Kunin 1999; Schaub, Jakober & Stauber 2013) which complicates inference about survival estimated with the classical CJS model (Cilimburg *et al.* 2002; Marshall *et al.* 2004; Zimmerman, Gutierrez & LaHaye 2007). The spatial CJS model does not suffer from these pitfalls. This allows obtaining more accurate estimates of survival from capture–recapture data collected in an arbitrarily defined study area. Comparison of survival of species between study areas or between groups of individuals with differential dispersal (e.g. ages, sexes) now becomes possible. In addition, direct inference about dispersal is possible. The study of dispersal is complicated due to the definition of the study area (Koenig, Van Vuren & Hooge 1996; Schneider 2003; Fujiwara *et al.* 2006; van Noordwijk 2011) in a similar way as is survival, and our model provides a solution for this problem.

Any dispersal movement is Markovian in space, and therefore, emigration from the study area cannot be characterized by the types of emigration that was used traditionally in capture–recapture studies (i.e. random temporary, non-random temporary and permanent). These traditional types of emigration assume that all individuals have the same probability to emigrate and that absent individuals have the same probability to return to the study population (0 when emigration is permanent, >0 when emigration is temporary). These assumptions do not apply for dispersing individuals, since the probability to emigrate from the study area and to return to the study area depends on their actual location and is thus not the same in all individuals. The spatial CJS model allows the estimation of an emigration rate, which is the proportion of individuals that is outside the study area at a given location. This parameter may be of interest if the model is used in population studies (e.g. in integrated population models; Schaub & Abadi 2011).

Long-term capture–recapture studies can be subject to an extension or contraction of the study area in the course of the study. If not accounted for, this induces spurious changes of apparent survival. Hitherto, one typically only used a restricted part of the data to mitigate the problem, either from the area that was covered during the whole study period or from a restricted period of time when the study area did not change. That is, some valuable information had to be discarded. With our new model, the complete data can be used and improved estimates of survival are obtained. Correction due to the change of the study area is done with the state-space **A** that is variable over time.

There are also limits in the application of the model. If dispersal movements become large compared to the size of the study area and capture probability is low, survival and recapture probabilities as well as dispersal variance are all positively biased when the data are analysed with the spatial CJS model. While the bias in recapture probability and dispersal variance can be substantial, it is moderate in survival. If dispersal distance is large and recapture probability low, the sample of the observed dispersal events gets more biased towards short dispersal movements and the sample size generally declines. The biases likely originate from the difficulty to estimate correctly the dispersal variance from a biased, small sample of observed dispersal distances, which is a common problem when dispersal kernels are estimated (Fujiwara *et al.* 2006). We are not aware of any model to reduce the biases in this situation. The only solution seem to enlarge the study area, to add a supplemental study area where some of the dispersing individuals arrive (Marshall *et al.* 2004; Gilroy *et al.* 2012) or to include additional information about dispersal from other data (e.g. radio- or satellite-tracking). The MSE of the estimates generally became larger in irregularly shaped study areas compared to an equally sized square study area due to smaller number of recaptures. As a matter of design, it makes sense to pick study areas that minimize the edge per unit study area in order to maximize observed dispersal sample size. Therefore, squares or circles should be desirable.

Our simulation study has shown that the spatial CJS model works well, if the assumptions of the model are fulfilled. Perhaps the most critical assumptions that may not be fully fulfilled in given data sets are that the dispersal distribution is wrongly specified, that the recapture probability is not spatially homogenous and/or that the demarcation of the study area is fuzzy. While there is potential to relax some of these assumptions by further development of the model (see below), a future simulation study should find out how strongly the parameters of interest are affected, if assumptions are not fully met. Goodness-of-fit tests would help to detect violations of model assumptions, yet such tests need to be developed and evaluated first. Ergon & Garnder (2014) have suggested posterior predictive checks for the spatial open robust design model, and similar tests might be useful for the spatial CJS model.

The case study has shown sensitivity of the estimated survival probabilities to the choice of the dispersal model (Fig. 6). For accurate estimation of true survival, the choice of an appropriate dispersal model is important (see also Ergon &

Garnder 2014). We have used the DIC to choose which dispersal model was best supported by the data. The selection of the most appropriate dispersal model is likely to become more difficult with decreasing survival and recapture probabilities and with increasing dispersal. Goodness-of-fit testing for the dispersal model might be a desirable alternative for deciding which one performs in a satisfactory way.

The estimation of the dispersal kernel is a key component in dispersal studies. This is often challenging because the observed dispersal movements are censored due to the limited size of the study area (Koenig, Van Vuren & Hooge 1996; Schneider 2003) and because a fraction of dispersal events within the study area may remain undetected. Dispersal kernels are biased if censoring and imperfect detection are not accounted for (Skarpaas, Shea & Bullock 2005; Hirsch *et al.* 2012). Our model provides a solution to the problem of imperfect detection within the study area as we explicitly model the detection process. It also helps to overcome the problem of censoring. However, if censoring becomes too strong (dispersal distances very large relative to the size of the study area), our model is not successful anymore in correctly estimating dispersal and consequently survival. But it is hard to imagine that any model would succeed in obtaining meaningful estimates in this case.

Though there are some conceptual similarities, a main difference between our model to the model developed recently by Gilroy *et al.* (2012) is that, in the latter, estimation is performed in two steps. A spatial projection of dispersal distances around each capture location is estimated first from the observed dispersal events, and this information is used in a second step in the capture–recapture model to estimate true survival. In our model, the estimation is performed in one step. This is an advantage because it allows full assessment of the estimation errors and because the inclusion of covariates for survival, dispersal and recapture is straightforward. Thus, our fully model-based framework is more rigorous and flexible from a statistical point of view. Moreover, our model does not need to assume that all emigration is permanent as does the model of Gilroy *et al.* (2012).

There are several possibilities for the further development of our model. First, other dispersal models could be defined. Here we have modelled movements in *x* and *y* direction separately with two normal and *t* distributions. More flexibility could be achieved using mixture distributions, such that the movement in *x* and *y* direction is estimated with zero-inflated distributions (Ghosh, Mudkhopadhyay & Lu 2006; Ergon & Garnder 2014) or with mixtures of two or more normal or *t* distributions. The former would allow one to more flexibly model the proportion of individuals that do not disperse at all. An alternative option is to specify the dispersal model directly with a dispersal distance and a dispersal angle, as is commonplace in movement models (McClintock *et al.* 2012; Ergon & Garnder 2014). This might have the advantage of obtaining better inference about dispersal. For example, covariates could be related directly to the dispersal distance, which has a more immediate biological interpretation than when covariates are related to the distances moved in *x* and *y* direction. Secondly, spatially explicit information about habitat in the study area (and preferably also

beyond) could be used to model dispersal (Ovaskainen 2004; Royle *et al.* 2013a,b). If information about habitat beyond the study area is available, we expect that this additional information would improve the estimates of survival and dispersal. It may also open an avenue for relaxing the assumption of identical survival and dispersal behaviour inside and outside the study area. Thirdly, additional data that are informative about some parameters can easily be included in an integrated analysis (Schaub & Kéry 2012). Such additional data could be radio- or satellite-tracking data that are informative about dispersal or additional captures during the breeding season (Pollock 1982) that are informative about recapture probability. Particularly, when dispersal is strong and the observed dispersal distances are heavily censored, we expect that the inclusion of information about dispersal from tracking data could be a very powerful way to increase accuracy. Fourthly, models could be developed to relax the assumption of spatially homogenous recapture probabilities. A potential avenue is to grid the study area and to estimate grid-specific recapture probabilities. Presumably spatial random effects are necessary to render the parameters of such a model identifiable. Fifth, if the activity centre (i.e. nest location in our example) is not directly observable, but only locations of individuals within their home ranges are available, the activity centre has to be estimated. This seems possible, if several observations within the breeding season are available. Thus, it would require data sampling under the robust design, and the model becomes more similar to the model described by Ergon & Garnder (2014).

Further developments should also consider different sampling designs. The spatial CJS model introduced here assumes that encounters are possible everywhere within the study area, consistent with an 'area search' type of protocol (Royle & Young 2008) which is common in studies of birds and reptiles. This has the consequence that the study area is known and does not need to be estimated. Thus, the spatial CJS model is not suited for a sampling design where recaptures are only possible at fixed locations, such as in nest box or camera trapping studies, because there the sampling area is unknown.

In conclusion, the proposed model is a major step forward for solving the problem of the negative bias of survival from which most conventional capture–recapture models suffer. The additional information about capture location that is required by the model is often available in capture–recapture studies. Besides the estimation of true survival, the model also allows inference to be made about dispersal. Thus, it should open up many new possibilities for population analyses based on the data from marked individuals.

Acknowledgements

References

Barker, R.J. (1997) Joint modeling of live-recapture, tag-resight, and tag-recovery data. *Biometrics*, **53**, 666–677.

Brooks, S.P. & Gelman, A. (1998) Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, **7**, 434–455.

Burnham, K.P. (1993) *Marked Individuals in the Study of Bird Populations* (ed. J.D. Lebreton), pp. 199–213. Birkhäuser, Basel.

Cilimburg, A.B., Lindberg, M.S., Tewksbury, J.J. & Hejl, S.J. (2002) Effects of dispersal on survival probability of adult yellow warblers (*Dendroica petechia*). *The Auk*, **119**, 778–789.

Ergon, T. & Garnder, B. (2014) Separating mortality and emigration: modelling space use, dispersal and survival with robust design spatial-capture-recapture data. *Methods in Ecology and Evolution*, doi: 10.1111/2041-210X.12133

Fujiwara, M., Anderson, K.E., Neubert, M.G. & Caswell, H. (2006) On the estimation of dispersal kernels from individual mark-recapture data. *Environmental and Ecological Statistics*, **13**, 183–197.

Ghosh, S.K., Mudkhopadhyay, P. & Lu, J.-C. (2006) Bayesian analysis of zero-inflated regression models. *Journal of Statistical Planning and Inference*, **136**, 1360–1375.

Gilroy, J.J., Virzi, T., Boulton, R.L. & Lockwood, J.L. (2012) A new approach to the "apparent survival" problem: estimating true survival rates from mark-recapture studies. *Ecology*, **93**, 1509–1516.

Gimenez, O., Rossi, V., Choquet, R., Dehais, C., Doris, B., Varella, H., Vila, J.-P. & Pradel, R. (2007) State-space modelling of data on marked individuals. *Ecological Modelling*, **206**, 431–438.

Hirsch, B.T., Visser, M.D., Kays, R. & Jansen, P.A. (2012) Quantifying seed dispersal kernels from truncated seed-tracking data. *Methods in Ecology and Evolution*, **3**, 595–602.

Kéry, M. & Schaub, M. (2012) *Bayesian Population Analysis Using WinBUGS. A Hierarchical Perspective*, 1st edn. Academic Press, Waltham.

Koenig, W.D., Van Vuren, D. & Hooge, P.N. (1996) Detectability, philopatry, and the distribution of dispersal distances in vertebrates. *Trends in Ecology and Evolution*, **11**, 514–517.

Lebreton, J.D., Burnham, K.P., Clobert, J. & Anderson, D.R. (1992) Modeling survival and testing biological hypothesis using marked animals: a unified approach with case studies. *Ecological Monographs*, **62**, 67–118.

Lebreton, J.D., Nichols, J.D., Barker, R.J., Pradel, R. & Spendelow, J.A. (2009) Modeling individual animal histories with multistate capture–recapture models. *Advances in Ecological Research*, **41**, 87–173.

Marshall, M.R., Diefenbach, D.R., Wood, L.A. & Cooper, R.J. (2004) Annual survival estimation of migratory songbirds confounded by incomplete breeding site-fidelity: study designs that may help. *Animal Biodiversity and Conservation*, **27.1**, 59–72.

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

van Noordwijk, A.J. (2011) Measuring dispersal as distance-dependent recruitment rates: testing the performance of DDRR on simulated data. *Journal of Ornithology*, **152**, S239–S249.

Ovaskainen, O. (2004) Habitat-specific movement parameters estimated using mark-recapture data and a diffusion model. *Ecology*, **85**, 242–257.

Pasinelli, G., Müller, M., Schaub, M. & Jenni, L. (2007) Possible causes and consequences of philopatry and breeding dispersal in red-backed shrikes *Lanius collurio*. *Behavioral Ecology and Sociobiology*, **61**, 1061–1074.

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)* (eds K. Hornik, F. Leisch & A. Zeileis), pp. 1–10. Vienna, Austria. Available at http://www.ci.tuwien.ac.at/Conferences/DSC-2003/Proceedings/.

Pollock, K.H. (1982) A capture–recapture design robust to unequal probability of capture. *The Journal of Wildlife Management*, **46**, 752–757.

Powell, L.A., Conroy, M.J., Hines, J.E., Nichols, J.D. & Krementz, D.G. (2000) Simultaneous use of mark-recapture and radiotelemetry to estimate survival, movement, and capture rates. *The Journal of Wildlife Management*, **64**, 302–313.

R Development Core Team (2004) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna.

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

Royle, J.A. & Young, K.Y. (2008) A hierarchical model for spatial capture-recapture data. *Ecology*, **89**, 2281–2289.

Royle, J.A., Chandler, R.B., Gazenski, K.D. & Graves, T.A. (2013a) Spatial capture–recapture models for jointly estimating population density and landscape connectivity. *Ecology*, **94**, 287–294.

Royle, J.A., Chandler, R.B., Sollmann, R. & Gardner, B. (2013b) *Spatial Capture–Recapture*. Elsevier, Amsterdam.

Schaub, M. & Abadi, F. (2011) Integrated population models: a novel analysis framework for deeper insights into population dynamics. *Journal of Ornithology*, **152**(Suppl. 1), 227–237.

Schaub, M., Jakober, H. & Stauber, W. (2011) Demographic response to environmental variation in breeding, staging and non-breeding areas in a migratory passerine. *Oecologia*, **167**, 445–459.

Schaub, M., Jakober, H. & Stauber, W. (2013) Strong contribution of immigration to local population regulation: evidence from a migratory passerine. *Ecology*, **94**, 1828–1838.

Schaub, M. & Kéry, M. (2012) Combining information in hierarchical models improves inferences in population ecology and demographic population analyses. *Animal Conservation*, **15**, 125–126.

Schaub, M., Gimenez, O., Schmidt, B.R. & Pradel, R. (2004) Estimating survival and temporary emigration in the multistate capture–recapture framework. *Ecology*, **85**, 2107–2113.

Schneider, C. (2003) The influence of spatial scale on quantifying insect dispersal: an analysis of butterfly data. *Ecological Entomology*, **28**, 252–256.

Skarpaas, O., Shea, K. & Bullock, J.M. (2005) Optimizing dispersal study design by Monte Carlo simulation. *Journal of Applied Ecology*, **42**, 731–739.

Thomas, C.D. & Kunin, W.E. (1999) The spatial structure of populations. *Journal of Animal Ecology*, **68**, 647–657.

Turchin, P. & Thoeny, W.T. (1993) Quantifying dispersal of southern pine beetles with mark-recapture experiments and a diffusion model. *Ecological Applications*, **3**, 187–198.

Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and Management of Animal Populations*, 1st edn. Academic Press, San Diego, California.

Zimmerman, G.S., Gutierrez, R.J. & LaHaye, W.S. (2007) Finite study areas and vital rates: sampling effects on estimates of spotted owl survival and population trends. *Journal of Applied Ecology*, **44**, 963–971.

Supporting Information

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

Appendix S1. R code to simulate data and analyze them with JAGS for square study areas.

Appendix S2. R code to simulate data and analyze them with JAGS for irregularly shaped study areas.

Appendix S3. Additional results from the simulation study using the square study area with 15 km side length.

Appendix S4. Description and additional results of the simulation study using an irregularly shaped study area.

Appendix S5. Design and results of the simulation study to investigate the effect of the size and the shape of the study area on the parameter estimates.