

Open Capture–Recapture Models with Heterogeneity: II. Jolly–Seber Model

Shirley Pledger,^{1,*} Kenneth H. Pollock,^{2,**} and James L. Norris^{3,***}

¹School of Mathematics, Statistics and Operations Research, Victoria University of Wellington, P.O. Box 600, Wellington, New Zealand

²Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695, U.S.A.

³Department of Mathematics and Computer Science, Wake Forest University, Winston-Salem, North Carolina 27109-7388, U.S.A.

**email:* shirley.pledger@vuw.ac.nz

***email:* pollock@stat.ncsu.edu

****email:* norris@mthesc.wfu.edu

SUMMARY. Estimation of abundance is important in both open and closed population capture–recapture analysis, but unmodeled heterogeneity of capture probability leads to negative bias in abundance estimates. This article defines and develops a suite of open population capture–recapture models using finite mixtures to model heterogeneity of capture and survival probabilities. Model comparisons and parameter estimation use likelihood-based methods. A real example is analyzed, and simulations are used to check the main features of the heterogeneous models, especially the quality of estimation of abundance, survival, recruitment, and turnover. The two major advances in this article are the provision of realistic abundance estimates that take account of heterogeneity of capture, and an appraisal of the amount of overestimation of survival arising from conditioning on the first capture when heterogeneity of survival is present.

KEY WORDS: Abundance; Capture–recapture; Finite mixture model; Heterogeneity; Jolly–Seber; Maximum likelihood; Non-parametric MLE; Open populations; Schwarz–Arnason.

1. Introduction

Open population capture–recapture models are used for populations with imperfect detection, to estimate important parameters such as population size, recruitment, and survival. These parameters in turn give information on turnover and population growth, with all parameters essential for understanding the dynamics of the population.

The Jolly–Seber (JS) model (Jolly, 1965; Seber 1965) provides these estimates for the simple sampling scheme, in which single samples are taken with a large spacing over time. Schwarz and Arnason (1996) used an entry parameter idea from Crosbie and Manly (1985) to provide a fully likelihood-based version of JS; we label this model JSSA.

Model JS assumes homogeneity of capture and survival probabilities over the whole population. If heterogeneity is present, this model gives misleadingly precise underestimates of population size, a problem that is more marked with heterogeneity of capture than with heterogeneity of survival (Carothers, 1973, 1979; Pollock et al., 1990). Earlier methods of adjusting JS to allow for heterogeneity of capture were not likelihood-based (see, e.g., Hwang and Chao, 1995; Pledger and Efford, 1998). This article proposes likelihood-based models, which provide maximum likelihood estimates of the parameters, confidence intervals, likelihood ratio tests, Akaike's information criterion (AIC; Akaike, 1973) for model comparisons, and profile likelihood intervals where asymmetry of the likelihood surface renders symmetric confidence intervals unrealistic (Cormack, 1992).

Finite mixtures (nonparametric MLE) have been used to model heterogeneity in closed populations with some success at bias correction in population size (Norris and Pollock, 1995, 1996; Pledger, 2000, 2005). They have also been used for open populations (Pledger, Pollock, and Norris, 2003), extending the Cormack–Jolly–Seber model (CJS; Cormack, 1964; Jolly, 1965; Seber, 1965; Lebreton et al., 1992). However, abundance is not estimated by CJS and is not discussed in Pledger et al. (2003). Our new likelihood-based extension to JSSA uses finite mixtures to (i) reduce bias in abundance estimates by modeling heterogeneity of capture, and (ii) reduce bias in survival estimates by modeling the uncaught animals as well as those caught. These developments are needed for realistic population modeling, where abundance and survival estimates are important inputs, and biases in these estimates have substantial consequences. The models have identifiability within the class of finite mixture models.

Section 2 gives the background data information and assumptions, Section 3 specifies the models, and Section 4 gives model comparison and estimation methods. A real example is shown in Section 5 and appraisals by simulation are in Section 6. Section 7 has discussion and conclusions.

2. Assumptions and Notation

The assumptions are those of Pledger et al. (2003), augmented with Schwarz and Arnason's (1996) superpopulation of N animals, each of which is present and available for capture on at least one sampling occasion. Proportions $\beta_0, \beta_1, \dots, \beta_{K-1}$ of

the N animals enter the population and are first available for capture at times 1, 2, \dots K , respectively ($\sum \beta_j = 1$).

There are C animal classes with membership unknown (a latent effect); each animal comes independently from class c with probability π_c ($\sum \pi_c = 1$). An animal from class c , if present at sample j , has probability p_{jc} of capture in sample j , and probability ϕ_{jc} of survival to the next sample (with ϕ_{Kc} assumed to be zero). There are D distinct animals seen, with $x_{ij} = 1$ if animal i is caught in sample j , otherwise $x_{ij} = 0$. Animal i 's capture history, CH_i , is the row vector \mathbf{x}_i , with first and last captures at t_i , l_i respectively. The $N - D$ uncaught animals each have capture history $\text{CH}_0 = \mathbf{0}$, a K -vector of zeros. There are n_h animals with capture history h .

3. The Models

Our models extend those of Pledger et al. (2003), using a multinomial allocation of the N animals to their capture histories. Writing the parameters N , β , π , p , and ϕ as a parameter vector θ , the likelihood for the “full model” is

$$L(\theta \mid \text{data}) = \frac{N!}{(N - D)! \prod_h n_h!} \times \prod_{i=1}^D L_i \times L_0^{N-D}, \quad (1)$$

where h indexes the different observed capture histories. The individual likelihoods are found by summing over the classes and all feasible birth and death times (b and d , respectively, samples when first and last available for capture):

$$L_i = P(\text{CH}_i) = \sum_{b=1}^{f_i} \sum_{d=\ell_i}^K \sum_{c=1}^C \left[\pi_c \beta_{b-1} \left(\prod_{j=b}^{d-1} \phi_{jc} \right) (1 - \phi_{dc}) \times \left\{ \prod_{j=b}^d p_{jc}^{x_{ij}} (1 - p_{jc})^{1-x_{ij}} \right\} \right], \quad (2)$$

(where the empty product $\prod_{j=b}^{d-1} \phi_{jc} = 1$ if $b = d$). Similarly for an uncaught animal,

$$L_0 = P(\text{CH}_0) = \sum_{b=1}^K \sum_{d=b}^K \sum_{c=1}^C \left[\pi_c \beta_{b-1} \left(\prod_{j=b}^{d-1} \phi_{jc} \right) (1 - \phi_{dc}) \times \left\{ \prod_{j=b}^d (1 - p_{jc}) \right\} \right]. \quad (3)$$

We label the full model $\{\beta(t), \phi(t \times h_C), p(t \times h_C)\}$, to indicate that β depends on time (sample), while ϕ and p both allow interactively for time effects and heterogeneity (via a finite mixture model with C classes).

For computational effectiveness and the provision of more appropriate confidence intervals, we reparameterize the model, expressing the parameters ϕ_{jc} and p_{jc} on a logit scale, and N on a log scale. Using τ for time and η for heterogeneity,

$$\text{logit}(p_{jc}) = \log \left(\frac{p_{jc}}{1 - p_{jc}} \right) = \mu + \tau_j + \eta_c + (\tau\eta)_{jc} \quad (4)$$

with constraints $\sum \tau_j = 0$, $\sum \eta_c = 0$, and each row and column of $(\tau\eta)_{jc}$ adding to 0. Similarly $\text{logit}(\phi_{jc})$ may be modeled with main effects and interaction.

Models with interactive time or heterogeneity effects may be more complicated than necessary, and have too many pa-

rameters for successful model fitting. A simpler model for p_{jc} with additive effects of time and class has $\text{logit}(p_{jc}) = \mu + \tau_j + \eta_c$ with constraints $\sum \tau_j = 0$ and $\sum \eta_c = 0$. Similarly survival could be modeled as additive on the logit scale. Further simplifications have capture and/or survival probabilities dependent only on time (e.g., $\text{logit}(p_{jc}) = \mu + \tau_j$), only on individual heterogeneity ($\text{logit}(p_{jc}) = \mu + \eta_c$), or constant over time and animals ($\text{logit}(p_{jc}) = \mu$), with similar simplified versions of ϕ_{jc} .

We use the notation $\phi(t \times h)$, $\phi(t + h)$, $\phi(t)$, $\phi(h)$, and $\phi(\cdot)$ for interactive, additive, time, heterogeneity, and constant effects, respectively, on survival, and a similar notation for these effects on capture probability. Hence $\{\beta(t), \phi(t \times h_C), p(t + h_C)\}$ denotes a model with C classes, time effects on the entry parameters β , interactive effects on survival, and additive effects on capture probability. With some algebra, the likelihood of model $\{\beta(t), \phi(t), p(t)\}$ reduces to that of JSSA (Schwarz and Arnason, 1996). Any model with ϕ constant over time ($\phi(h)$ or $\phi(\cdot)$), has ϕ adjusted to some standard time unit, e.g., annual ϕ . If class c has constant annual survival ϕ_c and the interval between samples j and $j + 1$ is t_j years, $\phi_{jc} = \phi_c^{t_j}$ is used in the likelihood formula.

In many of these models there is some parameter redundancy. Details of numbers of redundant parameters and our computational solutions are in Web Appendix 1. With enough data, all models are feasible except the full model.

3.1 Incorporating Extra Information

It is usual to allow for losses on capture (Pradel, 1996; Schwarz and Arnason, 1996), where some animals are known to die or are removed at sample j . We see such “deaths” as just one of four processes that are easily incorporated into our individual-based models. These are natural or unnatural death (e.g., death on capture, or removal), and natural or unnatural arrival (e.g., translocation from another population). No extra parameters (e.g., probability of loss on capture) are needed, as we simply modify L_i (equation 2) for that animal.

Unnatural death, when a previously seen individual i is captured at sample ℓ_i , but dies due to capture or is removed, has the term $(1 - \phi_{dc})$ in L_i omitted, and d fixed at ℓ_i with no summation over possible death times. In this way, the animal contributes information to the likelihood up to and including its final capture, but right-censoring prevents it from contributing to the estimation of ϕ (which measures only natural death or emigration).

Natural death, when a previously seen individual i has a known natural death or emigration between samples j^* and $j^* + 1$, has d fixed at j^* (with no sum over possible death times), but the term $(1 - \phi_{dc})$ is retained as the death or departure was natural.

Unnatural arrival, with animal i injected into the population at sample f_i , has b set to f_i (with no summation over possible entry times), but β_{b-1} and p_{bc} terms are not included in L_i . In this way, the unnatural part of its history does not contribute to the parameter estimation (left-censoring), but after its placement in the population it starts contributing to survival and capture probability estimates. Note that by excluding β_{b-1} we are not counting this animal in the superpopulation N , which is seen as a “natural” superpopulation for the study area. These extra animals must be added in to

the abundance estimates later, especially if population models with density-dependent parameters are to be fitted subsequently, using the capture-recapture estimates.

Natural arrival, with individual i known to have entered the population naturally between samples $j^* - 1$ and j^* , has b set to j^* , and β_{j^*} retained.

3.2 Other Parameters

Some functions of the basic parameters N , β_j , π_c , and ϕ_{jc} are of particular biological interest. The following formulae come from expected values, and the derived parameters may be estimated by plugging in the basic parameter estimates.

3.2.1 Current population size N_j . In the JS model N_j , the population size at the time of sample j is a parameter of major importance for population monitoring and density-dependence modeling. The number in class c present at sample j is found from the difference equation

$$N_{j+1,c} = N_{jc} \phi_{jc} + \pi_c N \beta_j \quad \text{with} \quad N_{1,c} = \pi_c N \beta_0 \quad (5)$$

assuming class is independent of entry time. Then $\hat{N}_j = \sum_{c=1}^C \hat{N}_{jc}$. The likelihood may be reparameterized with terms N_j replacing N and β_j . This is useful in population monitoring if a submodel with constraints on N_j is required, for example to test if N_j is constant over time, or if it shows a linear trend over all the samples or over the most recent samples. Equation (5) is used to switch from parameters $\{N, \beta_j, \pi_c, \phi_{jc}\}$ to $\{N_j, \pi_c, \phi_{jc}\}$, and the reverse operation given $\{N_j, \pi_c, \phi_{jc}\}$ is the recursion: (i) initialize with $B_0 = N_1$ and for each c , $N_{1c} = \pi_c N_1$, (ii) for $j = 2, \dots, K$ let $B_{j-1} = N_j - \sum_c N_{j-1,c} \phi_{j-1,c}$ and then for each c let $N_{jc} = N_{j-1,c} \phi_{j-1,c} + \pi_c B_{j-1}$, (iii) let $N = \sum_{j=0}^{K-1} B_j$ and $\beta_j = \frac{B_j}{N}$.

3.2.2 Seniority γ_{jc} . Extending Pradel's (1996) definition of seniority to classes, the probability a class c animal alive at $j + 1$ was also alive at j is $\gamma_{j+1,c} = \frac{N_{jc} \phi_{jc}}{N_{j+1,c}}$, $j = 1, \dots, K - 1$.

3.2.3 Turnover, $T_{j,j+1}$. If ϕ is homogeneous we measure population turnover between samples j and $j + 1$ as a weighted average of the proportion of N_j departing and the proportion of N_{j+1} arriving during the interval. Allowing for C classes of animal, we define turnover as

$$T_{j,j+1} = \frac{\sum_{c=1}^C N_{jc}(1 - \phi_{jc}) + \sum_{c=1}^C N_{j+1,c}(1 - \gamma_{j+1,c})}{\sum_{c=1}^C N_{jc} + \sum_{c=1}^C N_{j+1,c}}.$$

Other definitions of turnover are possible (see e.g., Boulenger et al., 1998). We note that our definition is equivalent to (1 - Sørensen's index between samples j and $j + 1$), with one representing a complete turnover of individuals and zero indicating no turnover. (Sørensen's index of association is used in community analysis to compare two samples for presence/absence of species.) Because of insufficient opportunities for earlier or later captures, the turnover estimates between the first two samples and between the last two samples have positive bias.

3.2.4 Structure of current population. The current population N_j may be partitioned into four types of animal, by cross-classifying them as seniors (present at the last sample) or new arrivals, and as stayers (present at the next sample) or those about to depart. We label senior stayers as

residents, senior nonstayers as retirees, new stayers as settlers, and new nonstayers as transients (a group including short-lived animals as well as animals on the move). Within class c , for $j = 2, \dots, K - 1$, the four estimates are given by: (i) Residents: ${}^+N_{jc}^+ = \gamma_{jc} N_{jc} \phi_{jc}$, (ii) Settlers: ${}^-N_{jc}^+ = (1 - \gamma_{jc}) N_{jc} \phi_{jc}$, (iii) Retirees: ${}^+N_{jc}^- = \gamma_{jc} N_{jc} (1 - \phi_{jc})$, and (iv) Transients: ${}^-N_{jc}^- = (1 - \gamma_{jc}) N_{jc} (1 - \phi_{jc})$. Summation over the classes gives the overall estimated numbers in the four groups, ${}^+N_j^+$, ${}^-N_j^+$, ${}^+N_j^-$, and ${}^-N_j^-$. A large proportion of residents indicates slow turnover in relation to the time intervals between samples.

3.2.5 Covariates. Covariates in time may be incorporated into this scheme. For example, a time effect in the probability of capture due to weather or varying search effort could be accounted for by modeling $\text{logit} p_{jc} = \mu + \eta_c + \gamma x_j$ where x_j is a relevant weather covariate, or represents search effort, at sample j , and the parameter γ is a logistic regression coefficient. Similarly survival may be modeled with time- or sample-dependent covariates.

We cannot model individual covariates, as these are not available for the uncaught animals.

3.2.6 Multiple Populations. The animals may come from two or more spatially, taxonomically, or sexually separate populations, which are groups with known membership (Lebreton et al., 1992; Schwarz and Arnason, 1996). Joint likelihoods over the groups are products of the individual likelihoods for each population, and groups may be compared, for example to see if two groups have similar survival patterns or trends in N_j over time.

4. Statistical Analysis

4.1 Exploratory Analysis

A range of plausible models is fitted, and compared using AIC (Akaike, 1973; Burnham and Anderson, 2002) or BIC. Mixture models do not satisfy regularity conditions normally required for AIC comparisons, but Burnham and Anderson found that if a mixture model fits with parameter estimates interior to the parameter space, it may be included in AIC comparisons.

4.2 Confirmatory Analysis

If a single appropriate model (the "complete model") has already been chosen, the likelihood may be used to estimate parameters and their standard errors, and to test for possible simplifications by likelihood ratio tests.

For interval estimates, we recommend either profile likelihood intervals (Cormack, 1992) or confidence intervals back-transformed from symmetric asymptotic intervals for $\log(N)$, $\text{logit}(\beta)$, $\text{logit}(\phi)$, or $\text{logit}(p)$. Where the Hessian of the optimization does not provide reasonable estimates of standard errors, bootstrap intervals may be used.

Likelihood-ratio tests (LRTs) can test for equality or certain patterns in the parameters. If the reduced model (assuming the null hypothesis) has fewer mixture classes than the complete model, it is a nonstandard LRT, and a bootstrap test (Norris and Pollock, 1996) or the methods in Self and Liang (1987) are used. Examples of LRTs follow. (i) A closure test, $H_0: \beta_0 = 1$ and all $\phi_{jc} = 1$. Acceptance of H_0 shows there is no evidence of births and deaths occurring. (ii) Comparing a complete model with $\phi(t \times h)$ with a reduced

Table 1

Possum data, $K = 9$ samples, $D = 270$ animals seen. Relative AIC values for 24 feasible models. Note the dichotomy of AIC values between models with and without heterogeneity of capture.

ϕ model	$p(\cdot)$	$p(t)$	$p(h_2)$	$p(t + h_2)$	$p(t \times h_2)$
$\phi(\cdot)$	39.0	34.9	13.1	4.1	8.2
$\phi(t)$	29.4	37.6	0.0	2.6	9.1
$\phi(h_2)$	43.0	30.2	15.0	5.7	7.7
$\phi(t + h_2)$	33.4	41.6	1.6	4.2	10.5
$\phi(t \times h_2)$	47.4	55.6	9.0	16.5	NA

model $\phi(t + h)$ to decide if the classes have varying or similar patterns of survival over time. (iii) A test comparing $p(t + h)$ with $p(t)$ shows if, after allowing for time effects, there is heterogeneity of capture. This is a nonstandard LRT.

4.3 Computing

We used full likelihood maximization from the **optim** command in **R** (R Development Core Team, 2007). The evaluation of the likelihood was speeded up by writing this function in **C++** and calling it from **R**. We used the default convergence criteria in **optim**, and always obtained convergence. Because of possible multimodalities in mixture model likelihood surfaces, we tried a range of starting points for each optimization.

5. Real Example

Murray Efford provided data from live trapping of the Australian brushtail possum (*Trichosurus vulpecula* Kerr) on a study grid in the Orongorongo Valley, near Wellington, New Zealand. There were 270 animals from $K = 9$ samples taken in February 1980–1988, and enough different capture histories were observed for the models to be identifiable (see Web Appendix 1).

Viewing the data as an “exploratory study,” 24 feasible models (using two classes) were fitted. A check of models with three classes in the finite mixture showed no improvement, so only two-class models are shown. The relative AIC (AIC – minimum AIC) results are in Table 1.

Selection of the model $\{\phi(t), p(h_2)\}$ accords well with other knowledge of this population. There is spatially induced heterogeneity of capture, as the traps are always set in the same location on a grid, and possums with a home range including a trap are more likely to be captured. Note that two other models had AIC quite close to the chosen model; model $\{\phi(t + h_2), p(h_2)\}$ with relative AIC only 1.6 suggests possible heterogeneity in ϕ , while model $\{\phi(t), p(t + h_2)\}$ with relative AIC 2.6 suggests possible time effects in p .

From the $p(t)$ column of Table 1, we note that an artificial restriction of p to time effects only causes selection of $\phi(h_2)$. The ϕ and p parameter estimates are correlated, so if heterogeneity is disallowed in p it appears in ϕ , which may be misleading. It is better to see the full picture, using all 24 feasible models.

Table 1 shows a clear dichotomy, with low relative AIC (< 17) for models with heterogeneity in p , and high relative AIC (> 29) for homogeneous- p models. The influence of heterogeneity of p on population size estimates is shown in

Figure 1, where confidence intervals for N are consistently higher and wider for models involving heterogeneity of p .

Figure 1 shows that for estimating N , the exact choice of model is not important, provided heterogeneity of p is included—both point estimates and confidence intervals match well over the best 14 models. (This feature carries over to the N_j estimates, not shown here.) If p is assumed to be homogeneous, there is an apparent 15–20% underestimation of N and N_j , matching results in Pledger and Efford (1998), where simulation and inverse prediction were used to correct bias in N using different data from the same possum population.

Alternatively, we may view the data set as a “confirmatory study.” Previous knowledge of this population (Efford, 1998; Pledger and Efford, 1998) suggests that $\{\beta(t), \phi(t), p(t + h_2)\}$ would be an appropriate model. (This was the third choice in the exploratory analysis, with AIC only 2.6 above that of the “best” model.) The chosen model gives $\hat{N} = 365$, with a 95% confidence interval (331, 419). Three hypothesis tests are of interest.

- (i) Are arrivals constant over time (apart from the initial β_0 which has an accumulation of previous arrivals over a few years)? $H_0: \beta_1 = \beta_2 = \dots = \beta_{K-1}$ is rejected at a 5% significance level (LRT, $\chi^2 = 16.6555$ on 6 d.f., $p = 0.0106$).
- (ii) Is annual survival constant over time? $H_0: \phi_1 = \phi_2 = \dots = \phi_{K-1}$ is rejected at a 5% level (LRT, $\chi^2 = 13.5259$ on 6 d.f., $p = 0.0354$).
- (iii) Is capture probability constant over time? Comparison with the model $\{\beta(t), \phi(t), p(h_2)\}$ gives an acceptance (LRT, $\chi^2 = 9.4469$ on 6 d.f., $p = 0.1500$).

The N_j estimates and their partitions are shown in the bar plot of Figure 2. The large proportion of residents (dark shading) is typical of a population with little turnover (high survival and low recruitment). The percentage turnover is shown between the bars, including the overestimates at the start and end.

Arrival β_j and survival ϕ_j estimates $\pm 1\text{SE}$, are shown in Figure 3. The lower survival ϕ_5 , which occurred during an unusually wet winter, appears to have been followed by increased recruitment. This is consistent with the known high densities, with occupancy of a limited number of home ranges, suggesting a population near the carrying capacity of the land.

6. Simulation and Appraisal

Simulation studies provided appraisals of our models and methods. Full details of these studies are in Web Appendix 2; a summary of our findings is presented here. Each simulation run generated capture histories for 100 populations of size $N = 400$, usually using parameters near to those estimated for the possum population. Unseen animals were removed from the generated capture matrix before analysis. Estimation of N was appraised using medians rather than means, as occasional unrealistically high N estimates can occur. This phenomenon, which has also been observed in closed population analyses (Norris and Pollock, 1996; Dorazio and Royle, 2003; Morgan and Ridout, 2008), occurs due to vagaries of the data from some generated populations. We used $\text{MAD} = \text{median}$

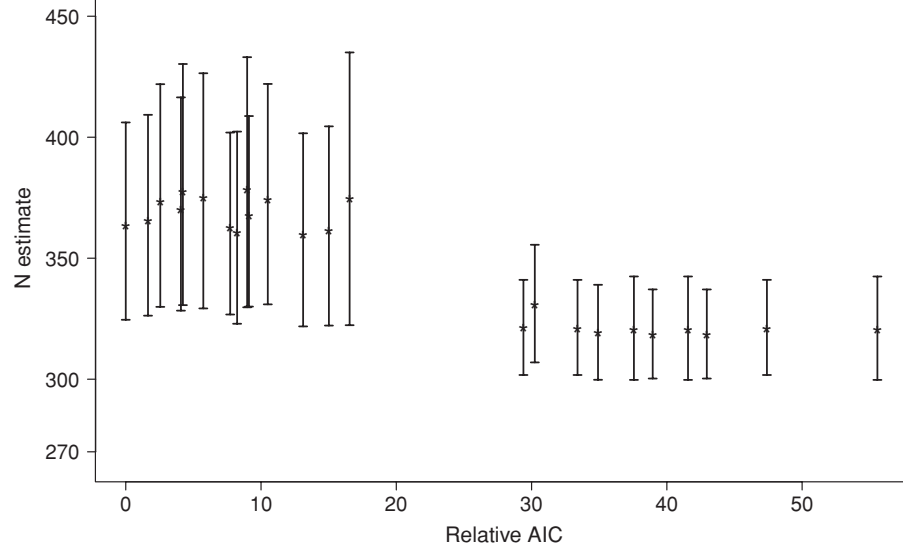


Figure 1. Possum data: 95% confidence intervals for superpopulation N versus relative AIC for the 24 feasible models. The intervals are back-transformed from asymptotic intervals for $\log(N)$. The 14 models with low AIC and high \hat{N} are precisely those with heterogeneity in p . There were 270 animals seen.

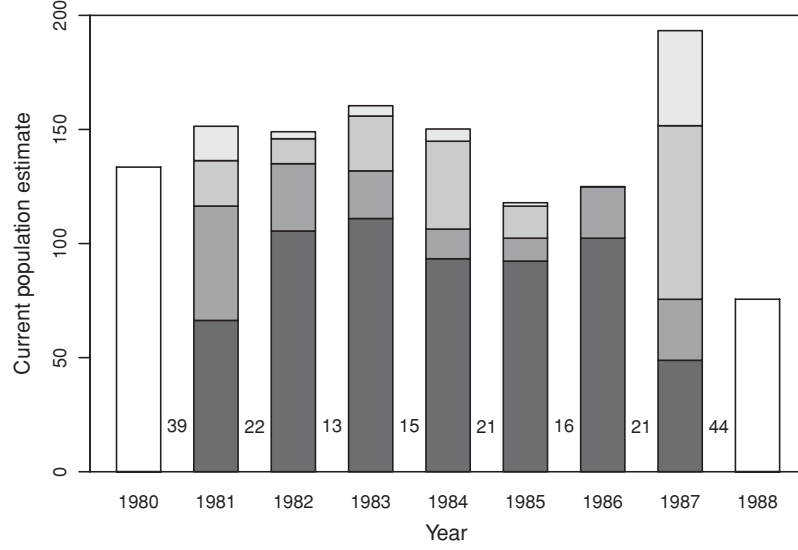


Figure 2. Barplot of \hat{N}_j over time for the possum data. The bars are partitioned into estimated numbers of (i) ${}^+N_j^+$ = residents = animals present before and after year j (dark shading), (ii) ${}^-N_j^+$ = settlers = new animals who will stay to next year (medium-dark), (iii) ${}^+N_j^-$ = retirees = senior animals about to die or depart (medium-light), and (iv) ${}^-N_j^-$ = transients = newly arrived animals about to die or depart (light shading). The first and last bars are not shaded, as only partial partitioning is possible. The percent turnover is shown between the bars. The first and last turnovers are overestimated, due to insufficient opportunities for capture at the beginning and end.

absolute deviation (of the set of 100 estimates from the input value).

6.1 Simulation Study for Exploratory Analysis

The first study evaluated exploratory analyses, checking the AIC model selection and point estimates of the superpopulation N from fitting the 24 feasible models (Section 3). Two types of generating models were tested, a two-group mixture and a beta (infinite) mixture, each with $K = 5$ or 10 samples.

Similar results were obtained for both types of generating model. With only $K = 5$ samples, there was no single clear choice of best model, although in 84% of populations a model with heterogeneous p was selected. All models with heterogeneity in p had $\text{MAD}(\hat{N})$ between 6 and 17, while all models with homogeneity of p had $\text{MAD}(\hat{N})$ between 47 and 56. The sharp distinction between the two types of models confirms the importance of allowing for heterogeneity of p when estimating abundance. An excessively high value of \hat{N} ($\hat{N} >$

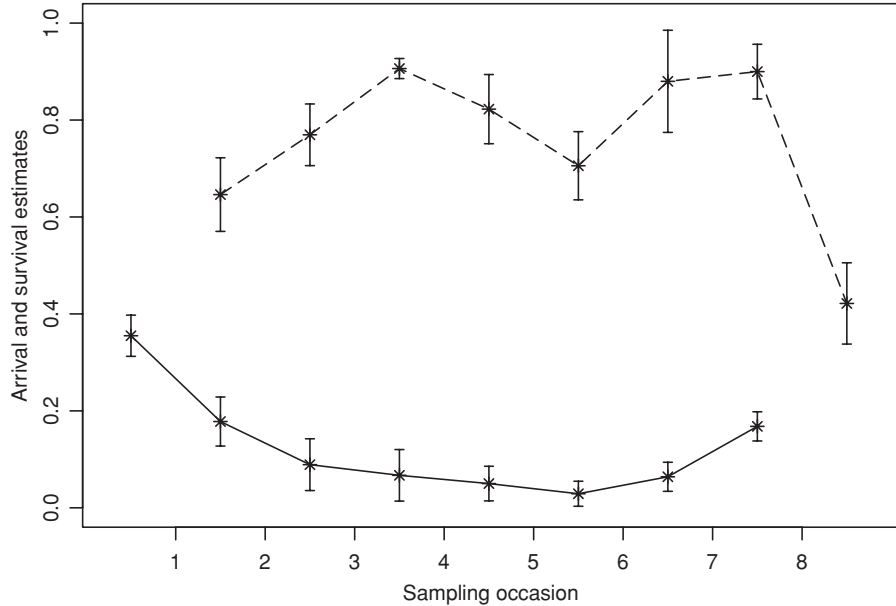


Figure 3. Possum data: estimated parameters for between-sample arrival (β_j , solid lines) and survival (ϕ_j , dashed lines) ± 1 standard error. Arrival before sample 1 is higher, having an accumulation of several years' animals, and the final survival estimate between samples 8 and 9 is the usual underestimate. Arrival following sample 8 is omitted due to near nonidentifiability. The heights of the curves are not comparable as the proportions do not have the same denominators, but the patterns may be compared. Lower survival between samples 5 and 6 appears to be followed by increased recruitment in the next two time intervals.

3D) occurred only once in 2400 model fits. With $K = 10$ samples, models with heterogeneous p were always selected, and there were no longer any excessively high N estimates (maximum $\hat{N} = 502$). All heterogeneous p models had $\text{MAD}(\hat{N})$ between 14 and 18, while all homogeneous p models had $\text{MAD}(\hat{N})$ between 34 and 45. Over both few and many sampling occasions, we see a clear split of $\text{MAD}(\hat{N})$ between homogeneous and heterogeneous models for p . This is similar to results from closed population models, which may be seen as a special case of our models under the restrictions $\beta_0 = 1$ and all $\phi = 1$. If abundance estimates are of interest, there is a strong case for checking if models with heterogeneous p are selected. Several heterogeneous models may provide similarly good estimates of N , with a flow-on effect to the estimates of N_j , population size at sample j (studied in more detail in the next simulations).

6.2 Simulation Studies for Confirmatory Analysis

Several simulation studies explored confirmatory analyses. Each had 100 populations, with superpopulation $N = 400$ and $K = 10$ samples. Appraisals used MAD, median relative bias (MRBias, %), and the coverage of nominal 95% confidence intervals (back-transformed from asymptotic symmetric interval for $\log N$ or $\log \phi$).

6.3 Influence of Heterogeneous p or ϕ on Abundance and Survival Estimates

Data generation with heterogeneity of capture confirmed underestimation of N with homogeneous models. The JS model had 64% coverage and MRBias -19.4% , while the heterogeneous model had 95% coverage, and MRBias -5.9% . Both models gave similar point estimates of survival, indicating lit-

tle effect of capture heterogeneity on ϕ . However, coverage of 89% from the JS model was improved to 93% with the heterogeneous model, which had larger standard errors. Data generated with heterogeneous ϕ showed little difference in \hat{N}_j when analyzed by JS or the correct model.

6.4 Influence of Assuming Two Classes for p When There Are Four Classes

Data were generated from four classes with capture probabilities 0.1, 0.2, 0.3, and 0.7, respectively, and analyzed by a two-class model. Coverages were 0.955 for abundance and also 0.955 for survival, with MRBias -2.64% and 0.22% respectively. It appears that with this much heterogeneity in p , two classes can adequately replace the four classes.

6.5 Influence of Assuming a Finite Mixture for p When the Mixture Is Infinite

The generating model had a beta distribution for capture probabilities, but data were analyzed using a two-class mixture. Coverages were 0.915 for abundance and 0.935 for survival, with MRBias -3.84% and 0.25% respectively. In this region of the parameter space, two classes appear to adequately model the beta distribution inputs.

6.5 Influence of Unmodeled Uncaught Animals

Populations were simulated with heterogeneity of survival, with 200 animals on constant $\phi = 0.9$, and 200 on $\phi = 0.55$. Analysis by the correct model had $\text{MAD}(\hat{\phi}) = 0.02$ and $\text{MRbias}(\hat{\phi}) = 0.16\%$. However, analysis by the correct model, but conditioning on first capture and excluding uncaught animals (as in Pledger et al., 2003), had $\text{MAD}(\hat{\phi}) = 0.05$ and $\text{MRbias}(\hat{\phi}) = 5.73\%$. This overestimation of ϕ is a result of

ignoring the uncaught animals, which are likely to be short-lived.

7. Discussion

The major innovation in this article is the combination of modeling the entry parameters (Crosbie and Manly, 1985; Schwarz and Arnason, 1996) with using finite mixtures for heterogeneity (Pledger et al., 2003), in order to reduce bias in abundance estimates. Real data and simulations confirmed the value of these models. A substantial extra advantage of this individual-based modeling is the easy incorporation of any available extra information about births and deaths.

Another major result from the new models is the evaluation of bias in survival estimates induced (in the presence of heterogeneity of survival) by models that condition on the first capture, and do not model the uncaught animals. Much single state and multistate modeling in capture-recapture uses this conditioning (CJS; Lebreton et al., 1992). Even if heterogeneity of survival is included in the model (as in Pledger et al., 2003), there is still overestimation of survival. If there is heterogeneity of survival, short-lived animals have lower probability of at least one capture, and models that ignore the uncaught animals will still overestimate survival. There are likely to be flow-on effects for other calculations, e.g., estimated lifespan of animals. The implications could be far reaching, suggesting that some effort should be made to check the amount of underestimation in the conditional models. (We note that this problem does not occur with another type of conditional modeling, found in Sanathanan, 1972; Huggins, 1989; and Fewster and Jupp, 2009. These models do allow for uncaught animals.)

We need not assume the animal classes actually exist, and if they do exist our estimates may not accurately reflect the true proportions and the within-class capture and survival probabilities (Lindsay, 1995). It has been found with closed population models that the exact details of the mixtures are often irrelevant when using mixtures to correct for bias in \hat{N} ; only with extreme heterogeneity does the nature of the mixture become important (Pledger, 2000, 2005). The mixtures are merely a device to incorporate heterogeneity into the model in order to reduce bias in the population abundance estimates, and to make profile likelihood interval coverage more accurate. A measure of heterogeneity is provided by mixture modeling (Dorazio and Royle, 2003; Pledger, 2005), and an exploration of its properties (e.g., stability, variability) would be a useful development.

Unidentifiability is an issue with both closed and open population models (Link, 2003), and a model for the heterogeneity is needed (Holzmann, Munk, and Zucchini, 2006; Holzmann and Munk, 2008; Mao, 2008). As detailed in Web Appendix 1, there is identifiability of the superpopulation N , and hence the abundance estimates over time, provided (i) a small number of redundant parameters are allowed for in the model fitting, and (ii) there is enough information in the data set.

In closed population capture-recapture, finite mixture models and models based on the beta distribution have been compared (Dorazio and Royle, 2003; Pledger, 2005) and combined (Morgan and Ridout, 2008). With moderate heterogeneity, finite mixtures and the (infinite) beta mixture often provide similar abundance estimates, both offering a

considerable improvement over models assuming homogeneity. With more extreme heterogeneity there may be discrepancies among heterogeneous models, with each approach having some scenarios in which it is better for bias reduction (Pledger, 2005). The number of parameters is of more importance than whether a finite or infinite mixture is used (Lindsay, 1995; Pledger, 2005). The beta distribution has two parameters and can approximate the mean and variance of the distribution of p , while a two-class mixture, with three parameters, can reflect skewness as well as mean and variance. Further simulations (unpublished) showed that the more parsimonious beta distribution is preferred if the skewness in the generating distribution happened to match that of the beta distribution, while an unmatched skewness tends to favor a two-class finite mixture. Skewness has been found to be important in heterogeneous extensions to the JS model (Pledger and Efford, 1998). Our simulations in Section 6 showed that analysis using finite mixtures still gave substantial reduction of bias in abundance estimates, even when the generating model used a beta distribution. We recognize that in some regions of the parameter space nonidentifiability can occur, and suggest the practical approach of simulating using parameters estimated from the real data set of interest (as we did), and trying a range of models to check if the results are comparable.

A Bayesian approach is possible for the problem of heterogeneity-induced bias in abundance estimation for open populations. Another approach could use random effects for the time variation; this would reduce the number of parameters, although likelihood evaluation would become more difficult.

8. Supplementary Materials

Web Appendices referenced in Sections 3, 5, 6, and 7 are available under the Paper Information link at the *Biometrics* website <http://www.biometrics.tibs.org>.

ACKNOWLEDGEMENTS

The authors thank Murray Efford for possum data, guidance on using C++, and many helpful suggestions, and Ray Brownrigg for computing help and advice.

REFERENCES

- Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, B. N. Petrov and F. Csaki (eds), 267–281. Budapest: Akademiai Kiado.
- Boulinier, T., Nichols, J. D., Sauer, J. R., Hines, J. E., and Pollock, K. H. (1998). Estimating species richness: The importance of heterogeneity in species detectability. *Ecology* **79**, 1018–1028.
- Burnham, K. P. and Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd edition. New York: Springer-Verlag.
- Carothers, A. D. (1973). The effects of unequal catchability on Jolly-Seber estimates. *Biometrics* **29**, 79–100.
- Carothers, A. D. (1979). Quantifying unequal catchability and its effect on survival estimates in an actual population. *Journal of Animal Ecology* **48**, 863–869.
- Cormack, R. M. (1964). Estimates of survival from the sighting of marked animals. *Biometrika* **51**, 429–438.

- Cormack, R. M. (1992). Interval estimation for mark-recapture studies of closed populations. *Biometrics* **48**, 567–576.
- Crosbie, S. F. and Manly, B. F. J. (1985). Parsimonious modeling of capture-mark-recapture studies. *Biometrics* **41**, 385–398.
- Dorazio, R. M. and Royle, J. A. (2003). Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics* **59**, 351–364.
- Efford, M. G. (1998). Demographic consequences of sex-biased dispersal in a population of brushtail possums. *Journal of Animal Ecology* **67**, 503–517.
- Fewster, R. M. and Jupp, P. E. (2009). Inference on population size in binomial detectability models. To appear in *Biometrika*.
- Holzmann, H. and Munk, A. (2008). Rejoinder to “On the nonidentifiability of population sizes.” *Biometrics* **64**, 979–981.
- Holzmann, H., Munk, A., and Zucchini, W. (2006). On identifiability in capture-recapture models. *Biometrics* **62**, 934–939.
- Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika* **76**, 133–140.
- Hwang, W.-D. and Chao, A. (1995). Quantifying the effects of unequal catchabilities on Jolly-Seber estimates via sample coverage. *Biometrics* **51**, 128–141.
- Jolly, G. M. (1965). Explicit estimates from capture-recapture data with both death and immigration—stochastic model. *Biometrika* **52**, 225–247.
- Lebreton, J.-D., Burnham, K. P., Clobert, J., and Anderson, D. R. (1992). Modelling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs* **62**, 67–118.
- Lindsay B. G. (1995). *Mixture Models: Theory, Geometry and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics **5**. Institute of Mathematical Statistics, Hayward, California.
- Link, W. A. (2003). Nonidentifiability of population size from capture-recapture data with heterogeneous capture probabilities. *Biometrics* **59**, 1123–1130.
- Mao, C. X. (2008). On the nonidentifiability of population sizes. *Biometrics* **64**, 977–979.
- Morgan, B. J. T. and Ridout, M. S. (2008). A new mixture model for capture heterogeneity. *Applied Statistics* **57**, 433–446.
- Norris, J. L. III and Pollock, K. H. (1995). A capture-recapture model with heterogeneity and behavioural response. *Environmental and Ecological Statistics* **2**, 305–313.
- Norris, J. L. and Pollock, K. H. (1996). Nonparametric MLE under two closed capture-recapture models with heterogeneity. *Biometrics* **52**, 639–649.
- Pledger, S. (2000). Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics* **56**, 434–442.
- Pledger, S. (2005). The performance of mixture models in heterogeneous closed population capture-recapture. *Biometrics* **61**, 868–876.
- Pledger, S. and Efford, M. G. (1998). Correction of bias due to heterogeneous capture probability in capture-recapture studies of open populations. *Biometrics* **54**, 888–898.
- Pledger, S., Pollock, K. H., and Norris J. L. (2003). Open capture-recapture models with heterogeneity: I. Cormack-Jolly-Seber model. *Biometrics* **59**, 786–794.
- Pollock, K. H., Nichols, J. D., Brownie, C., and Hines, J. E. (1990). Statistical inference for capture-recapture experiments. *Wildlife Monographs* **107**, 3–97.
- Pradel, R. (1996). Utilization of capture-mark-recapture for the study of recruitment and population growth. *Biometrics* **52**, 703–709.
- R Development Core Team. (2007). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Sanathanan, L. (1972). Estimation of the size of a multinomial population. *Annals of Mathematical Statistics* **43**, 142–152.
- Schwarz, C. J. and Arnason, A. N. (1996). A general methodology for the analysis of capture-recapture experiments in open populations. *Biometrics* **52**, 860–873.
- Seber, G. A. F. (1965). A note on the multiple-recapture census. *Biometrika* **52**, 249–259.
- Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610.

Received September 2008. Revised September 2009.

Accepted September 2009.