

The use of auxiliary variables in capture-recapture modelling: an overview

KENNETH H. POLLOCK, *Department of Statistics, North Carolina State University, Raleigh NC 27695-8203, USA*

ABSTRACT *I review the use of auxiliary variables in capture-recapture models for estimation of demographic parameters (e.g. capture probability, population size, survival probability, and recruitment, emigration and immigration numbers). I focus on what has been done in current research and what still needs to be done. Typically in the literature, covariate modelling has made capture and survival probabilities functions of covariates, but there are good reasons also to make other parameters functions of covariates as well. The types of covariates considered include environmental covariates that may vary by occasion but are constant over animals, and individual animal covariates that are usually assumed constant over time. I also discuss the difficulties of using time-dependent individual animal covariates and some possible solutions. Covariates are usually assumed to be measured without error, and that may not be realistic.*

For closed populations, one approach to modelling heterogeneity in capture probabilities uses observable individual covariates and is thus related to the primary purpose of this paper. The now standard Huggins-Alho approach conditions on the captured animals and then uses a generalized Horvitz-Thompson estimator to estimate population size. This approach has the advantage of simplicity in that one does not have to specify a distribution for the covariates, and the disadvantage is that it does not use the full likelihood to estimate population size. Alternately one could specify a distribution for the covariates and implement a full likelihood approach to inference to estimate the capture function, the covariate probability distribution, and the population size.

The general Jolly-Seber open model enables one to estimate capture probability, population sizes, survival rates, and birth numbers. Much of the focus on modelling covariates in program MARK has been for survival and capture probability in the Cormack-Jolly-Seber model and its generalizations (including tag-return models). These models condition on the number of animals marked and released. A related, but distinct,

Correspondence: Kenneth H. Pollock, Department of Statistics, North Carolina State University, Raleigh NC 27695-8203, USA. E-mail: pollock@stat.ncsu.edu

topic is radio telemetry survival modelling that typically uses a modified Kaplan-Meier method and Cox proportional hazards model for auxiliary variables. Recently there has been an emphasis on integration of recruitment in the likelihood, and research on how to implement covariate modelling for recruitment and perhaps population size is needed. The combined open and closed ‘robust’ design model can also benefit from covariate modelling and some important options have already been implemented into MARK. Many models are usually fitted to one data set. This has necessitated development of model selection criteria based on the AIC (Akaike Information Criteria) and the alternative of averaging over reasonable models. The special problems of estimating over-dispersion when covariates are included in the model and then adjusting for over-dispersion in model selection could benefit from further research.

1 Introduction

Here I briefly review the use of auxiliary variables in capture-recapture models as they apply to the estimation of demographic parameters (e.g. population size, survival, recruitment, emigration, and immigration) for wild animal populations. Typically in the literature covariate modelling has made capture and survival probabilities functions of covariates, but it is logical to make other parameters functions of covariates as well. There are two very important reasons to model covariates. First it enables more parsimonious parameterizations and thus precision of all parameter estimates is increased. Second there may be inherent ecological importance in understanding the nature of the relationships between the parameters and specific environmental and individual animal variables.

The types of covariates considered include group or environmental covariates that may vary by occasion but are constant over animals and individual animal covariates that are usually assumed to be constant over time. I discuss the difficulties of using time-dependent individual animal covariates and some possible solutions. Also covariates are usually assumed to be measured without error and that may not be realistic.

While much research is being conducted, there are still enormous gaps in our knowledge of how to use covariates. In this paper I review what has been done and, where possible, suggest some alternative new approaches. Section 2 reviews closed models; Section 3 open models; Section 4 radio telemetry survival models; Section 5 combined open and closed models; Section 6 the special difficulties with time-dependent individual covariates. I conclude the paper with my views on other fruitful current and future research directions.

2 Closed models

2.1 Overview

I begin with a very brief review of closed capture-recapture models based on Pollock (1991, 2000) and other sources. Other important reviews are Schwarz & Seber (1999) and Chao (2001). The most basic model is the Lincoln-Petersen model for a closed population of size N . There is one sample of size M , where all animals are marked, and a second later one of size $n = (m + u)$ where both marked (m) and unmarked (u) animals are captured. Simple intuition suggests equating the sample proportion marked with the population proportion marked $m/n = M/N$. Thus our estimate is

$$\hat{N} = Mn/m$$

This estimator is maximum likelihood under a model with the following assumptions (Seber, 1982):

- (i) the population is closed to additions and deletions;
- (ii) all animals are equally likely to be captured in each sample; and
- (iii) marks are not lost and not overlooked by the observer.

The second assumption may be violated in two ways:

- (i) heterogeneity which occurs when different animals have fundamentally different capture probabilities; and
- (ii) trap response when the probability of capture depends on the animal's prior capture history.

Often one samples the population more than twice and each time every unmarked animal caught is uniquely marked, previously marked animals have their captures recorded, and usually all animals are released back into the population. This more extensive sampling enables sophisticated modelling that permits unequal catchability due to heterogeneity and trap response. Otis *et al.* (1978) consider a set of eight models where capture probabilities vary due to time, heterogeneity, and trap response in all possible combinations. They also provided a computer program, CAPTURE, to compute the estimates and select between models.

The heterogeneity models that use a distribution of capture probabilities have caused theoretical difficulty for statisticians. An early ad hoc approach based on the 'jackknife' method (Burnham & Overton, 1978; Pollock & Otto, 1983) proved quite useful. Later Chao and her colleagues, in a series of papers, used a method based on sample coverage (Chao *et al.*, 1992; Lee & Chao, 1994; Chao, 2001). Some have suggested use of log linear models (Fienberg, 1972; Cormack, 1989; Agresti, 1994). Becker, Yip, and others have suggested use of martingales (Becker, 1984; Yip, 1989, 1991; Lloyd & Yip, 1991; Yip & Fong, 1993; Lloyd, 1994). Maximum likelihood estimation where the heterogeneity is modelled as a finite mixture distribution (usually with two or three support points) (Norris & Pollock, 1995, 1996; Pledger, 2000) is a recent development. We will return to this later as it leads naturally to modelling covariates to explain some of the heterogeneity in capture probabilities.

Another difficulty has been to find a good method of model selection for this series of closed models. The original approach in Otis *et al.* (1978) does not work well. I suspect that the finite mixture approaches will aid in solving this problem (Pledger, 2000). Model averaging may also be used (Burnham & Anderson, 1998; Stanley & Burnham, 1998).

2.2 Auxiliary variables

2.2.1 Environmental covariates.

Pollock *et al.* (1984) suggested linear logistic modelling of capture probabilities using observed environmental and other auxiliary variables, such as effort, which only varied by sampling occasion and they gave some examples. Their approach is very straightforward and is now widely used to model capture probabilities in both closed and open models. The model is conditional on the values of the variables observed, so that no attempt is made to model the distribution of the variables.

2.2.2 Individual animal covariates. The situation is more complex when individual animal auxiliary variables are used and I now discuss this in some detail. I consider the case where the covariate is time-invariant first, as it is much simpler to model statistically. Linear logistic modelling of capture probabilities, using individual auxiliary variables, may have the purpose of accounting for heterogeneity although some inherent heterogeneity may still remain due to other unobservable variables (Pollock *et al.*, 1984; Huggins, 1989, 1991; Alho, 1990; Borchers *et al.*, 1998; Pledger, 2000). In addition, the modelling should increase the precision of population size estimates and satisfy the principle of parsimony (Burnham & Anderson, 1998).

The paper by Pollock *et al.* (1984) is now only of historical interest in this context. They noted that, for individual animal covariates, one can only observe the covariates on the captured animals, but their solution based on homogeneous subgroups of animals was ad hoc.

There are two sensible modern approaches to modelling individual covariates. The first of these suggested independently by Huggins (1989) and Alho (1990) is to condition on the observed animals and then use a generalized Horvitz-Thompson estimator to estimate population size. This approach has the advantage of simplicity, in that one does not have to specify a distribution for the covariates, and the disadvantage that it does not use the full likelihood to estimate population size.

The second approach is to specify a distribution for the covariates and implement a full likelihood approach to inference where one has to estimate the parameters of the capture function, $p(x)$, which is the probability of capture for an animal conditional on its covariate x , the covariate probability distribution, $\pi(x)$, and the population size N . Typically $p(x)$ is modelled as a linear logistic (logit) function of the covariate so that

$$\ln [p(x)/(1 - p(x))] = \alpha + \beta x$$

This approach is suggested by Borchers *et al.* (1998) for the more general problem of multiple observer line transect models which can be viewed as a combination of line transects and closed capture-recapture models. It is different from the Huggins-Alho approach in that the parameters of $\pi(x)$ have to be estimated in addition to the parameters of $p(x)$ and N .

2.3 Statistical model development for individual covariates

2.3.1 Full likelihood. For simplicity I present the full likelihood for a k -sample capture-recapture study with one covariate x (having distribution $\pi(x)$ which is, for example, Normal with parameters μ and σ) and capture functions $p_1(x)$, $p_2(x), \dots, p_k(x)$ having different intercepts (α) but the same slope (β). A much more parsimonious sub-model would be to put all intercepts equal. However, the results developed here are much more general.

The development of the general likelihood equation follows that of Alpizar & Pollock (1996) and Borchers *et al.* (1998) for the more general problem of combining line transects and capture-recapture.

$$L_f = L_1 L_2$$

where

$$L_1 = \left[\frac{N!}{n!(N-n)!} \right] p.^n (1-p.)^n$$

and

$$L_2 = \prod_{i=1}^n \prod_h \left[\frac{p_h(x_i) \pi(x_i)}{p.} \right]$$

with

$$p. = \sum_h \int p_h(x) \pi(x) dx$$

Note that L_1 is binomial and that $p.$ is the probability of catching any animal at least once. Also note that $p_h(x)$ is the probability of capture for that history h , conditional on x , so that for the history captured only in times 1 and 2 it would be

$$p_1(x)p_2(x)(1-p_3(x)) \dots (1-p_k(x))$$

and so on. The summation is over all distinct h where an animal is captured at least once. In L_2 the product is over all animals captured at least once and the particular $p_h(x)$ is evaluated at the value x_i for animal i . Note that the problem of unobserved covariates does not arise because they are integrated out, making this a true likelihood.

The parameter estimates for $(\alpha_1, \alpha_2, \dots, \alpha_k, \beta, \mu, \sigma)$ and hence $p.$ would come from the second component of the likelihood. The parameter N could be estimated from the moment equation (Sanathanan, 1972) based on the first component of the likelihood.

$$\hat{N}_f = \frac{n.}{\hat{p}.}$$

or

$$\hat{N}_f = \frac{n.}{\sum_h \int p_h(x) \pi(x) dx.}$$

The first expression is a ubiquitous equation that occurs for all closed capture-recapture maximum likelihood estimators of N . The trick is to derive the probability of capturing any animal at least once for a particular model ($p.$). It is a function of the parameters in the model and has to be estimated.

There are important questions that need to be addressed before the full likelihood approach could be implemented and actual estimates computed. The expression $p.$ involves some very complex integration problems. Also there are many parameters to estimate so that in some cases the model may be non-identifiable. For instance, in the example illustrated it may be necessary to assume that all the intercepts (α) are equal unless k is very large. On the other hand, if these problems can be overcome, then a full maximum likelihood approach might have the advantage of better precision for the population size estimator than the conditional approach.

2.3.2 Conditional Huggins-Alho likelihood. Here the conditional likelihood is based on Huggins (1989) but I present it in a slightly modified form to contrast it to the full likelihood approach. The likelihood (conditional on the observed covariates) is

obviously only defined for captured animals because otherwise the covariates are unobserved! I also present the Horvitz-Thompson estimator of N suggested by Huggins (1989).

$$L_c = \prod_{i=1}^{n_c} \prod_h \left[\frac{p_h(x_i)}{\sum_h p_h(x_i)} \right]$$

$$\hat{N}_c = \sum_i^{n_c} \left[\frac{1}{\sum_h \hat{p}_h(x_i)} \right]$$

Note that the full likelihood and conditional likelihood estimators of N will only be algebraically equal in the trivial case where the capture probabilities do not depend on the covariate (x). One problem with this estimator is that sometimes the capture history probabilities in the denominator can be small and the estimator can become very unstable.

I postpone discussion of the situation when individual covariates are time-dependent until Section 3.2.1 because they are also relevant to the open models discussed next. The fact that the covariates are only observed on the occasions when the animals are captured creates special problems that are still largely ignored in the literature.

3 Open populations

3.1 Brief overview

3.1.1 General Jolly-Seber model. Often capture-recapture studies last a long time rendering the closed models impractical. Therefore there has been a need for the development of models that allow for additions (recruits and immigrants) and deletions (deaths and emigrants). The first general stochastic open model was developed independently by Jolly (1965) and Seber (1965). Their model, which requires equal catchability and equal survival rates of all animals at each sampling time, enables one to estimate population sizes, survival rates, and birth numbers for almost all samples. Detailed treatments are in Seber (1982) and Pollock *et al.* (1990). In the Jolly-Seber model survival and capture probabilities are estimated from the component of the likelihood that involves recaptures of marked animals. (This is now known as the Cormack-Jolly-Seber model.) Then traditionally these estimators are used to define moment equations to estimate population size and birth numbers using the component of the likelihood relating marked and unmarked animals. We now give an intuitive description of the estimators.

Given the number of marked animals, M_i , is known at each sampling occasion, then population size for period i can be determined using the Petersen estimator.

$$\frac{m_i}{n_i} \cong \frac{M_i}{N_i}$$

$$\hat{N}_i = \frac{\hat{M}_i n_i}{m_i} \quad (i = 2, \dots, k - 1)$$

Here, n_i represents the number of animals captured at each sampling occasion, m_i of which are marked. However, because deletions are now possible, in order to estimate abundance (and other vital rates), one must estimate M_i .

The survival rate estimator is obtained from the ratio of marked animals present at time $i + 1$ to time i ,

$$\hat{\phi}_i = \frac{\hat{M}_{i+1}}{\hat{M}_i + (R_i - m_i)} \quad (i = 1, \dots, k - 2)$$

where the term, $R_i - m_i$, represents the number of newly marked animals released at time i . The survival estimator does not distinguish between losses due to death and due to permanent emigration, without more information.

To estimate birth rates, the difference in population size at time $i + 1$ and time i is determined, by accounting for deaths due to natural mortality $(1 - \phi_i)$ and capture mortality $(n_i - R_i)$,

$$\hat{B}_i = \hat{N}_{i+1} - \hat{\phi}_i(\hat{N}_i - (n_i - R_i)) \quad (i = 2, \dots, k - 2)$$

The birth rate estimator cannot distinguish between individuals entering the population due to recruitment and immigration. However, in the case of recruitment, it may be possible to separate the individuals entering on the basis of size, sexual maturity, etc.

Capture probability, p_i , is estimated as the ratio of marked animals caught at time i to the number present in the population at time i .

$$\hat{p}_i = \frac{m_i}{\hat{M}_i} = \frac{n_i}{\hat{N}_i} \quad (i = 2, \dots, k - 1)$$

All of the above estimators rely on the knowledge (or estimation) of the marked population size, M_i . The marked population size is estimated by equating the future recapture rates of two distinct groups of animals in the population at time i , those that are marked (both previously and currently) but not seen at time i ($M_i - m_i$), and those that are seen and released at time i (R_i). Under the assumption of equal capture rates and equal survival rates of individuals, the future recapture rates of these two groups should be equivalent. Thus, if z_i and r_i are the members of the $M_i - m_i$ and R_i which are captured again later (at least once),

$$\frac{z_i}{M_i - m_i} \approx \frac{r_i}{R_i} \quad (i = 2, \dots, k - 1)$$

$$\hat{M}_i = m_i + \frac{R_i z_i}{r_i}$$

These closed-form estimators are useful for understanding the structure and assumptions of the Jolly-Seber model. However, in practice there is the need for the fitting of many sub-models of this general model. The sub-models could be simply simplifications where survival and/or capture rates are made constant over time or generalizations where survival and/or capture rates are made functions of covariates. In either case, parameters can no longer be estimated explicitly and complex computer programs are needed. In the next section, I focus on the Cormack-Jolly-Seber model for survival and capture rate estimation and then return to the more general Jolly-Seber model which includes recruitment estimation as well.

3.1.2 Cormack-Jolly-Seber model (survival modelling). There has been much research on modelling survival rates (and capture probabilities) allowing for multiple strata (e.g. sex, age, location). An important reference is Lebreton *et al.*

(1992). An interesting development has been the shift to fitting large numbers of models that has necessitated development of model selection criteria based on the AIC (Akaike Information Criteria). The alternative approach of averaging over some reasonable models is also being considered. The recent book by Burnham & Anderson (1998) on model selection is very important to this research.

Related models based on band recoveries of exploited animals (Brownie *et al.*, 1985) have been developed and are very widely used to analyse band return data on migratory waterfowl and other animals. These models led to important work on the compensatory versus additive mortality hypotheses for hunting mortality. Recently these models have begun to be used in fisheries studies in a slightly different form using instantaneous rates of fishing and natural mortality (Hoenig *et al.*, 1998a, 1998b).

A driving force in this recent research has been the availability of excellent software. The most notable is Program MARK (White & Burnham, 1999) that makes the fitting of many related models and choice among them straightforward. This program will be considered in more detail later when we discuss auxiliary variables.

Bayesian approaches to these complex survival models for capture-recapture and tag return models and their combination are now beginning to be used (Vounatsou & Smith, 1995; Dupuis, 1995; Brooks *et al.*, 2001). This is due to the development of efficient computer-based sampling methods to estimate posterior distributions using Markov Chain Monte Carlo (MCMC) and the BUGS software (Gilks *et al.*, 1996; Spiegelhalter *et al.*, 1999; Chen *et al.*, 2000).

3.1.3 Other models. Early survival modelling work considered fixed strata like sex or strata where the transition to the next stage was automatic (i.e. as in successive age classes). Recent works have looked at transitions between stages or states in a probabilistic manner. For example, animals may be marked in two different locations and, as well as surviving between periods, they may move to another location. Stage-structured models allow the estimation of both survival and movement probabilities. Some important papers that outline important theoretical results and provide interesting examples are Hestbeck *et al.* (1991), Nichols *et al.* (1992), Brownie *et al.* (1993), and Schwarz *et al.* (1993). MSSURVIV was written by Jim Hines to aid in these analyses (Brownie *et al.*, 1993). Research in this important area is continuing (e.g. Lebreton & Pradel 2001).

Combining data sources to improve precision of survival estimates and allow separation of apparent survival into true survival and fidelity rates has also been a source of much recent work. Two very important papers are Burnham (1993) and Barker (1997). Current work includes Nasution *et al.* (2001a, 2001b) and others.

Other recent work has emphasized integration of recruitment in the likelihood (e.g. Schwarz & Arnason, 1996; Pradel, 1996). Related work allows direct estimation of population change (λ) (Nichols & Hines, 2001).

3.2 Auxiliary variables

3.2.1 Survival and capture probability modelling. Modelling survival as a function of covariates goes back at least to 1979 when North & Morgan (1979) used winter temperature to model survival of herons. There have been many examples since then and recently a variety of software packages has been developed. A very important reference is Lebreton *et al.* (1992) who gave a variety of examples using

Program SURGE. Smith *et al.* (1994) developed Program SURPH and gave examples. POPAN-5 (Arnason & Schwarz, 1999) allows modelling of survival and capture probabilities as a function of group but not individual covariates.

Program MARK allows one to model both survival and capture probabilities as functions of both environmental and individual covariates (White & Burnham 1999; White *et al.*, 2000; Cooch, 2000a, 2000b; Franklin, 2000) for capture-recapture and tag return data and combinations of them. Cooch (2000b) and Franklin (2000) give very useful basic examples that show some of the many features of this extremely versatile and comprehensive program with covariates.

Analyses are conditional on the observed covariate values (i.e. the covariate distribution is not specified). A link function links the linear model specified in the design matrix with the survival, recapture, reporting rate, and fidelity probabilities specified in the parameter index matrix. Several link functions are available:

- Sin
- Logit (typically used when modelling covariates)
- Loglog
- Complementary loglog
- Log
- Identity

The Logit is the most common link used when modelling covariates and recall, for example, that the Logit $[\phi(x)] = \ln [\phi(x)/(1 - \phi(x))] = \alpha + \beta x$.

When computing the log likelihood function the following conceptual equation is used:

$$\begin{aligned} \text{Log } L \propto & \sum_h \text{Log (Pr (Observe encounter history } h)) \\ & \times (\text{No. animals with encounter history } h) \end{aligned}$$

For environmental covariates the basic unit is the encounter history with the summation over all capture histories h where an animal is captured at least once.

To illustrate some of the computational issues involved when covariates are modelled consider a 6-period live recapture study, with apparent survival, a logit linear function of a covariate such as air temperature at each sampling time, initial weight of each animal, or weight of each animal at each sampling time. One possible observed history is 101111 with

$$\begin{aligned} \text{Prob (Observe this encounter history)} \\ = \phi_1(x)(1 - p_2) \phi_2(x)p_3 \phi_3(x)p_4 \phi_4(x)p_5 \phi_5(x)p_6 \end{aligned}$$

Recall that in the Cormack-Jolly-Seber Model we are conditioning on marked animals released and we do not have to worry about modelling unmarked animals. This makes covariate modelling reasonably simple, at least conceptually. This is one of the simplest histories to write down, but the principle is the same for any of the capture histories.

For an environmental covariate such as air temperature at each sampling time, all animals with this or any other history have the same encounter probability, and hence they can be grouped. This greatly speeds up the computations. For an individual fixed covariate, such as initial weight of each animal, each animal has to have its encounter history probability computed separately because each animal has a distinct covariate value. This involves more intense computation.

Time-dependent individual covariates, such as weight of the animal at each sampling time, create special difficulties and cannot be computed in the current version of MARK (or any other program). Environmental and fixed individual animal covariates have the same x for all five survival probabilities (i.e. $\phi(x)$ uses same x for all periods). However, for time-dependent individual animal covariates we would have the encounter probability,

$$\phi_1(x_1)(1 - p_2)\phi_2(x_2)p_3\phi_3(x_3)p_4\phi_4(x_4)p_5\phi_5(x_5)p_6$$

with each animal in each time period having a distinct x value. Now x_2 is not observed because the animal is not captured at time 2 (note that x_1, x_3, x_4, x_5 are observed). We shall return to this later in Section 6. We note that the special case of a time-dependent covariate which changes in a known way (i.e. age) can be handled easily in existing software, because even when the animal is not captured, the value of the covariate is known.

Remember the analyses are conditional on the marked animals released and the observed values of covariates, and so the analyses are nice and simple, at least conceptually, for environmental and fixed individual animal covariates. One can standardize the covariates by subtracting off the mean and dividing by the standard deviation to avoid scaling problems. This is recommended when using MARK. One can add terms to allow for quadratic effects, or interactions between two covariates and so on. There are dangers of over-fitting and getting non-identifiable or near non-identifiable models for 'naive' users. It is crucial to look at model diagnostics and convergence very carefully before considering model parameter output. One also needs to use the model selection criteria, AIC, or its modifications to try and keep the final model chosen as parsimonious as possible. An alternative is to come up with a small set of potential models and use model averaging (Burnham & Anderson, 1998).

Model selection and precision for estimates under a given model depend for validity on assuming that the model fits the data well. When there are covariates then it is difficult to assess goodness of fit, because one can no longer use the traditional chi-square goodness of fit test used for capture-recapture models (see for example Seber, 1982). In addition, if the model is not fitting well, then the AIC and the standard errors of parameter estimates should be adjusted based on the estimate of the over dispersion parameter c (see Burnham & Anderson, 1998). Typically c is estimated as deviance/degrees of freedom (McCullagh & Nelder, 1989). White (2001) in his discussion of this paper notes that \hat{c} as estimated in Program MARK using a parametric bootstrap procedure (White & Burnham, 1999) can be severely negatively biased and he includes some simulations to show its magnitude.

3.2.2 Multi-state models. Spendelov *et al.* (1995) modelled movement as a function of environmental covariates (i.e. colony size and distance between colonies) in multi-state models for a meta-population of roseate terns using program MSSURVIV. The basic multi-state model (Schwarz *et al.*, 1993) has been implemented in MARK so that individual and group covariates can be used for modelling survival and movement rates.

3.2.3 Recruitment and population change estimation. I think a lot more research work is needed here. Recruitment as defined by Schwarz & Arnason (1996) can

currently be modelled as a function of an environmental covariate in POPAN-5 (Arnason & Schwarz, 1999) but it is not possible to model individual covariates.

The seniority parameter of Pradel (1996) can be modelled as a function of individual covariates using MARK (White & Burnham, 1999). The seniority parameter (γ_{i+1}) is the analog of a survival parameter when a capture-recapture study is run backwards in time and measures $[\phi_i/(\phi_i + f_i)]$. Thus $(1 - \gamma_{i+1})$ is a measure of $[f_i/(\phi_i + f_i)]$ which is per capita recruitment rate (f_i) as a fraction of per capita recruitment rate plus survival rate (ϕ_i) (Franklin, 2000). However, I think biologists would probably be more interested in modelling f_i itself as a function of covariates.

4 Radio-telemetry survival modelling

Biologists began the use of radio tags as a method of studying animal movements. As the technology improved and the sample sizes increased, this approach has also been used to estimate survival rates. Early papers were by Trent & Rongstad (1974) and Heisey & Fuller (1985). Pollock *et al.* (1989) used and modified the Kaplan-Meier method that is widely used in medical applications of survival analysis for wildlife telemetry. There have been many papers in the wildlife journals using these methods. In some cases the Cox proportional hazards model for relating auxiliary variables to survival has been used (e.g. Riggs & Pollock, 1992).

When first proposed it was assumed that these models required that these animals had capture probabilities of 1 (they did allow for right censoring). While it is reasonable to make this assumption in some cases Pollock *et al.* (1995) formulated these models as a Jolly-Seber capture-recapture model where both live and newly dead animals could be detected with equal capture probability less than one. They showed that the staggered entry Kaplan-Meier estimator can still be used for survival estimation. This model has not been used very much, however, it would be feasible to model survival or capture rates as a function of environmental or fixed individual animal covariates using program SURVIV or program MARK.

5 Combined open and closed models

5.1 Overview

Traditionally long-term capture recapture data have been analyzed using the Jolly-Seber model or its variants. These models were discussed in detail in Section 3. A problem with this approach is that the assumption of equal catchability is often difficult to satisfy and bias in population size estimates often results although survival rate estimates are usually much less affected. In addition, in many studies there may be multiple capture periods close together in time with the possibilities of recaptures. In a Jolly-Seber formulation these events would all be pooled and short-term recaptures ignored. Pollock (1982) suggested a design with primary periods spaced far enough apart so that an open model is necessary. However, he suggested an addition where each primary period had several secondary periods close together. For example a study might have three monthly primary periods and then within each month five consecutive trapping days. He suggested that the within primary period captures and recaptures be used in a closed population model to estimate population size in a manner robust to unequal capture probabilities (i.e. use one of the 8 models in Program CAPTURE). Survival rate would be estimated

using the Cormack-Jolly-Seber open model on the primary period captures and recaptures (i.e. pool all the captures within each primary period). Birth numbers can also be estimated using the deterministic equations proposed by Jolly (1965) but with the population estimates coming from the closed models rather than the open model. Nichols & Pollock (1990) realized that the robust design could be used to separate *in situ* recruitment from immigration if young and adults were both tagged.

Kendall & Pollock (1992) explored this ‘robust’ design further using simulation. Kendall *et al.* (1995) built models that provided more efficient maximum likelihood estimators than Pollock’s ad hoc approach when the closed component did not include heterogeneity. Kendall & Nichols (1995) and Kendall *et al.* (1997) realized that the design had another important benefit. The problem of temporary emigration has plagued open capture-recapture models and they realized that if it is reasonable to assume that the temporary emigration did not occur between the secondary periods of a primary period then a solution was possible.

SURVIV (White, 1983) can be used to fit any multinomial model, but it is not easy to use when there are a large number of cells to specify. Kendall & Hines (1999) developed a ‘front end’ to SURVIV called RDSURVIV so that parameter estimates could be obtained for many sub-models including those allowing time variation and trap response of capture probabilities, and the modelling of temporary emigration of animals (i.e. the animals move out of the study area for a period and then return).

5.2 Auxiliary variables

Program MARK includes analysis based on the robust design as one of its many options (Kendall, 2000). It goes beyond the earlier program RDSURVIV in a very important way because it allows an option for the use of fixed individual covariates by implementing Huggins conditional likelihood (Huggins, 1989 and Section 2) within each primary period. Recently Yip *et al.* (2001) also considered the modelling of covariates for the robust design using Huggins model as the basis for their inference.

I believe that the full likelihood approach, discussed earlier in Section 2, could also be implemented for the robust design. One drawback of the full likelihood for a single closed population is that there are large numbers of parameters to specify. However, in the robust design it might be possible to specify equality of parameters across primary periods, achieving some parsimony and improved precision for parameter estimation.

6 Time dependent individual covariates

The special problems of using time-dependent individual animal covariates, previously discussed in Section 3.2.1, should be investigated thoroughly. I have not been able to find any publications on this topic, although I believe that researchers have begun to think about it. The key issues here are that the values of the covariate are only observable when an animal is captured and that a comprehensive analysis would require information on the distribution of the covariate and how it changes over time. I suggest some possible approaches that could perhaps apply to either open or closed population studies.

Lebreton & Pradel (2001) state that multi-stratum models constitute a canonical

capture-recapture model for individual categorical covariates changing over time. If we are dealing with a continuous covariate changing over time, then one approach would be to approximate it as a series of discrete categories with transitions between them (Nichols pers. com.). Nichols *et al.* (1992) essentially followed this approach when they analysed a meadow vole capture-recapture data set with weights categorized into discrete weight states. The obvious advantage of this ad hoc approach is that one does not need to specify how animals change in weight or the distribution of weight within the population of animals. The disadvantage would be that, if a full model could be specified, then more precise estimates of parameters would result.

A full likelihood approach could be attempted and this would involve integrating out the unobserved covariates. However, it would be necessary to specify the distribution of the covariate and how it was changing over time. Another approach based on the full likelihood would be Bayesian. The use of Bayesian methods is particularly suggested here by the need for dealing with unobserved covariates. Bayesian methods automatically integrate out unobserved random variables using numerical integration or Markov Chain Monte-Carlo sampling methods (MCMC).

One might also be able to carry out an analysis conditional on the covariates and view it as a missing data problem. If the change in the covariate were modelled over time, I suspect the EM algorithm (Dempster *et al.*, 1977; Little & Rubin, 1987) might be useful. Van Duesen (2001) has suggested use of the EM algorithm in the standard Cormack-Jolly-Seber model to speed up computation times for large data sets. He mentions individual covariates in his model formulation but his example does not include them.

7 Discussion and future research

7.1 Discussion

Statisticians may think that the primary purpose of using covariates is to get a more robust model and avoid bias due to heterogeneity (in the closed models). Biologists, however, may be drawn to covariate modelling because they want to explore important ecological hypotheses on the relationship of survival or recruitment parameters to auxiliary variables such as animal weight. White (2001) notes, however, that cause and effect relationships cannot be established without experimental manipulation of the system. Therefore opportunities to incorporate covariates into designed, manipulative experiments should be considered wherever practically feasible.

The perennial statistical inference question of the trade off between model reality and parameter estimator precision becomes very complex when covariates are involved. (If full likelihood methods are used there will be even larger numbers of parameters involved.) Therefore model selection methods and the related topic of model averaging are very important because, when covariates are involved, there will be very large numbers of models to choose from.

7.2 Some possible research directions

7.2.1 Closed models. The closed population models are interesting in their own right, but also, in addition, because they are a key component of the 'robust design' where open and closed models are combined in one overall model. There is a need

for much more research on these closed models. I suggest some possibly fruitful avenues now.

While the conditional, Huggins-Alho approach is now widely used, it can be unstable. An implementation of the full likelihood approach needs to be done and then a comparison with the conditional likelihood approach made. Borchers *et al.* (1998) suggested this initially.

Recent research by Pledger (2000) and others involve combining the observed conditional covariate approach of Huggins (1989) and Alho (1990) with the finite mixture distribution approach for dealing with inherent (or unexplainable) heterogeneity. Thus heterogeneity is modelled by observed covariates, plus an additional random component modelled by a finite point mixture. Theoretically this strikes me as important as heterogeneity is unlikely to ever be totally explained by covariates. In practice there could be serious problems of over parameterization unless capture probabilities and numbers of capture occasions are both high.

A combination of line transects and capture-recapture methods needs to be considered further. This could be a very useful practical combination in some settings (aerial and ship surveys of marine mammals; bird counts with multiple observers). However, modelling additional covariates beyond perpendicular distance present practical difficulties. While Borchers *et al.* (1998) present a complete likelihood including covariates and a conditional likelihood including covariates (analogous to the Huggins (1989) likelihood for capture-recapture discussed earlier), they are clearly uncertain about which is preferable. They note that simulations showed that the conditional approach had large positive bias for population size estimates, if detectability is low, while on the other hand they note that the full likelihood is very complex and requires specifications for distributions of covariates. Note that the distribution of perpendicular distance is uniform if the transect is placed randomly, and in that case there is no additional complexity. Of course if the transects were not random (i.e. along trails or roads) modelling the distribution of distance more generally could be very informative, although I would never advocate those kind of transects.

7.2.2 Open models. Goodness of fit, model checking diagnostics, and the estimation of over dispersion, in the presence of covariates, need to be considered further. White (2001) emphasized that he thought that this area was the most important for future research.

Not much attention has been given to modelling recruitment or population change as a function of covariates. The modelling of survival and movement in multi-state models could also benefit by more research. Robust design covariate modelling could be quite rich because of the possibility of imposing constraints across the primary periods. Many of the points made about the closed population models research directions also apply to the robust design.

7.2.3 General questions. Scaling of covariates needs more attention. Also the very important problem of individual animal time-dependent covariates, and the need for future research in that area have already been considered in Section 3.2.1 and Section 6.

Brooks *et al.* (2001) indicate that Bayesian methods for the analysis of individual covariates in capture-recapture and tag return models are being devised. This alternative to standard approaches used in Program MARK, for example, is likely

to receive a lot more attention in the future. Bayesian methods could be used for both closed and open models.

Measurement errors in the covariates need to be considered very carefully. It is well known that measurement errors can cause bias in standard linear and nonlinear regression problems (Carroll *et al.*, 1995). Various authors including Alpizar (1997) have investigated measurement errors in the distance variable in line transect sampling, but I have not seen measurement error in covariates addressed in the capture-recapture context. Alpizar (1997) showed that both the bias and the variance of the measurement error distribution are important and can cause bias in different directions. Thus, even if the measurement error is unbiased, it can still cause a serious bias in the covariate regression parameter estimates.

Acknowledgements

I thank Jim Nichols for helpful comments on modelling time-dependent covariates. I thank Dr. B. S. Brunhuber for typing and editorial assistance.

REFERENCES

- AGRESTI, A. (1994) Simple capture-recapture models permitting unequal catchability and variable sampling effort, *Biometrics*, **50**, pp. 494-500.
- ALPIZAR-JARA, R. & POLLOCK, K. H. (1996) A combination line transect and capture-recapture sampling model for multiple observers in aerial surveys, *Journal of Environmental and Ecological Statistics*, **3**, pp. 311-327.
- ALPIZAR-JARA, R. (1997) Assessing assumption violations in line transect sampling. Ph.D. Dissertation, Biomathematics Program, North Carolina State University, Raleigh, North Carolina, U.S.A.
- ALHO, J. M. (1990) Logistic regression in capture-recapture models, *Biometrics*, **46**, pp. 623-635.
- ARNASON, A. N. & SCHWARZ, C. J. (1999) Using POPAN-5 to analyse banding data, *Bird Study*, **46**, pp. S157-S168.
- BARKER, R. J. (1997) Joint modelling of live recapture, tag-resight and tag-recovery data, *Biometrics*, **53**, pp. 666-677.
- BECKER, N. G. (1984) Estimating population size from capture-recapture experiments in continuous time, *Australian Journal of Statistics*, **26**, pp. 1-7.
- BORCHERS, D. L., ZUCCHINI, W. & FEWSTER, R. M. (1998) Mark-recapture models for line transect surveys, *Biometrics*, **54**, pp. 1207-1220.
- BROOKS, S. P., CATCHPOLE, E. A. & MORGAN, B. J. T. (2001) Bayesian animal survival estimation. In: B. J. T. MORGAN & D. L. THOMSON (Eds) *Statistical Analysis of Data From Marked Bird Populations. Journal of Applied Statistics* (This volume).
- BROWNIE, C., ANDERSON, D. R., BURNHAM, K. P. & ROBSON, D. S. (1985), *Statistical Inference from Band-Recovery Data. A Handbook*, 2nd edn (Washington, DC, US Fish and Wildlife Service Resource Publication 156).
- BROWNIE, C., HINES, J. E., NICHOLS, J. D., POLLOCK, K. H., & HESTBECK, J. B. (1993) Capture-recapture studies for multiple strata including nonmarkovian transition probabilities, *Biometrics*, **49**, pp. 1173-1187.
- BURNHAM, K. P. & OVERTON, W. S. (1978) Estimation of the size of a closed population when capture probabilities vary among animals, *Biometrika*, **65**, pp.625-633.
- BURNHAM, K. P. (1993) A theory for combined analysis of ring recovery and recapture data. In: J.-D. LEBRETON & P. M. NORTH (Eds), *Marked Individuals in the Study of Bird Population*, pp. 199-213 (Basel, Switzerland, Birkhauser, Verlag).
- BURNHAM, K. P. & ANDERSON, D. R. (1998) *Model Selection and Inference: A Practical Information Theoretic Approach*, (New York, Springer-Verlag).
- CARROLL, R. J., RUPERT, D. & STEFANSKI, L. A. (1995) *Measurement Errors in Nonlinear Models*. Monographs in Statistics and Applied Probability, 63 (London, Chapman & Hall).
- CHAO, A., LEE, S.-M. & JENG, S. L. (1992) Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal, *Biometrics*, **48**, pp. 201-216.

- CHAO, A. (2001). An overview of closed capture-recapture models, *Journal of Agricultural, Biological, and Ecological Statistics*, 6, pp. 158-175.
- CHEN, M.-H., IBRAHIM, J. G. & SHAO, Q.-M. (2000). *Monte Carlo Methods in Bayesian Computation* (New York, Springer-Verlag).
- COOCH, E. (2000a) First steps with Program MARK: Linear models. Hungarian Symposium paper. MARK webpage reference.
- COOCH, E. (2000b) Standing out in a crowd: Individual Covariates. Chapter 11 of Online workbook for Program MARK.
- CORMACK, R. M. (1989) Loglinear models for capture-recapture, *Biometrics*, 45, pp. 395-413.
- DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. (1977) Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B*, 39, pp. 1-38.
- DUPUIS, J. A. (1995) Bayesian estimation of movement and survival probabilities from capture-recapture data, *Biometrika*, 82, pp. 761-772.
- FIENBERG, S. E. (1972) The multiple-recapture census for closed populations and incomplete contingency tables, *Biometrika*, 59, pp. 591-603.
- FRANKLIN, A. B. (2000) Exploring ecological relationships in survival and estimating rates of population change using program MARK. Hungarian Symposium paper MARK webpage reference.
- GILKS, W. R., RICHARDSON, S. & SPIEGELHALTER, D. J. (1996) *Markov Chain Monte Carlo in Practice* (London, Chapman & Hall).
- HEISEY, D. M. & FULLER, T. K. (1985) Evaluation of survival and cause specific mortality rates using telemetry data, *Journal of Wildlife Management*, 49, pp. 668-674.
- HESTBECK, J. G., NICHOLS, J. D. & MALECKI, R. A. (1991) Estimates of movement and site fidelity using mark-resight data of wintering Canada Geese, *Ecology*, 72, pp. 523-533.
- HOENIG, J. M., BARROWMAN, N. J., HEARN, W. S. & POLLOCK, K. H. (1998a) Multi-year tagging studies incorporating fishing effort data, *Canadian Journal of Fisheries and Aquatic Sciences*, 55, pp. 1466-1476.
- HOENIG, J. M., BARROWMAN, N. J., POLLOCK, K. H., BROOKS, E. N., HEARN, W. S. & POLACHECK, T. (1998b) Models for tagging data that allow for incomplete mixing of newly tagged animals, *Canadian Journal of Fisheries and Aquatic Sciences*, 55, pp. 1477-1483.
- HUGGINS, R. M. (1989) On the statistical analysis of capture experiments, *Biometrika*, 76, pp. 133-140.
- HUGGINS, R. M. (1991) Some practical aspects of a conditional likelihood approach to capture experiments, *Biometrics*, 47, pp. 725-732.
- JOLLY, G. M. (1965) Explicit estimates from capture-recapture data with both death and immigration—stochastic model, *Biometrika*, 52, pp. 225-247.
- KENDALL, W. L. & POLLOCK, K. H. (1992) The robust design in capture-recapture studies: a review and evaluation by Monte Carlo simulation. In: D. R. MCCULLOUGH & R. H. BARRETT (Eds), *Wildlife 2001: Populations*, pp. 31-43 (New York, Elsevier Applied Science).
- KENDALL, W. L., POLLOCK, K. H. & BROWNIE, C. (1995) A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design, *Biometrics*, 51, pp. 293-308.
- KENDALL, W. L. & NICHOLS, J. D. (1995) On the use of secondary capture-recapture samples to estimate temporary emigration and breeding proportions, *Journal of Applied Statistics*, 22, pp. 751-762.
- KENDALL, W. L., NICHOLS, J. D. & HINES, J. E. (1997) Estimating temporary emigration using capture-recapture data with Pollock's robust design, *Ecology*, 78, pp. 563-578.
- KENDALL, W. L. & HINES, J. E. (1999) Program RDSURVIV: an estimation tool for capture-recapture data collected under Pollock's robust design, *Bird Study*, 46, pp. S32-S38.
- KENDALL, W. L. (2000) The robust design for capture-recapture studies: analysis using program MARK. Hungarian Symposium paper. MARK webpage reference.
- LEBRETON, J.-D., BURNHAM, K. P., CLOBERT, J. & ANDERSON, D. R. (1992) Modelling survival and testing biological hypotheses using marked animals: a unified approach with case studies, *Ecological Monographs*, 62, pp. 67-118.
- LEBRETON, J.-D. & PRADEL, R. (2001) Multi-stratum capture-recapture models: modelling incomplete capture histories. In: B. J. T. MORGAN & D. L. THOMSON (Eds) *Statistical Analysis of Data From Marked Bird Populations*. *Journal of Applied Statistics*, (This volume).
- LEE, S.-M. & CHAO, A. (1994) Estimating population size via sample coverage for closed capture-recapture models, *Biometrics*, 50, pp. 88-97.
- LITTLE, R. J. A. & RUBIN, D. B. (1987) *Statistical Analysis with Missing Data* (New York, Wiley).
- LLOYD, C. J. & YIP, P. (1991) A unification of inference from capture-recapture studies through martingale estimating equations. In: V. P. GODAMBE (Ed), *Estimating Functions*, pp. 65-88 (Oxford, Oxford University Press).
- LLOYD, C. J. (1994) Efficiency of martingale methods in recapture studies, *Biometrika*, 81, pp. 305-315.

- MCCULLAGH, P. & NELDER, J. A. (1989) *Generalized Linear Models*, 2nd edn (New York, Chapman & Hall).
- NASUTION, M., BROWNIE, C., POLLOCK, K. H. & BENNETTS, R. E. (2001a) Joint analysis of resighting and radio-telemetry data, *Journal of Agricultural, Biological, and Agricultural Statistics*, (In Press).
- NASUTION, M., BROWNIE, C. & POLLOCK, K. H. (2001b) Optimal design of resighting and radio-telemetry studies. In: B. J. T. MORGAN & D. L. THOMSON (Eds) *Statistical Analysis of Data From Marked Bird Populations. Journal of Applied Statistics*, (This conference).
- NICHOLS, J. D. & POLLOCK, K. H. (1990) Separate estimation of recruitment from emigration versus in situ reproduction using Pollock's robust design, *Ecology*, **71**, pp. 21-27.
- NICHOLS, J. D., SAUER, J. R., POLLOCK, K. H. & HESTBECK, J. B. (1992) Estimating transition probabilities for stage based population projection matrices using capture-recapture data, *Ecology*, **73**, pp. 306-312.
- NICHOLS, J. D. & HINES, J. E. (2001) Approaches for the direct estimation of lambda and demographic contributions to lambda using capture-recapture data. In: B. J. T. MORGAN & D. L. THOMSON (Eds) *Statistical Analysis of Data From Marked Bird Populations. Journal of Applied Statistics*, (This conference).
- NORRIS, J. L. & POLLOCK, K. H. (1995) Capture-recapture model MRI: Bivariate MLE of capture-recapture distribution and resulting model selection, *Journal of Environmental and Ecological Statistics*, **2**, pp. 305-313.
- NORRIS, J. L. & POLLOCK, K. H. (1996) Nonparametric maximum likelihood estimators for population size under two closed capture-recapture models with heterogeneity, *Biometrics*, **52**, pp. 639-649.
- NORTH, P. M. & MORGAN, B. J. T. (1979) Modelling heron survival using weather data, *Biometrics*, **35**, pp. 667-681.
- OTIS, D. L., BURNHAM, K. P., WHITE, G. C. & ANDERSON, D. R. (1978) Statistical inference from capture data on closed animal populations, *Wildlife Monographs*, **62**, pp. 1-135.
- PLEDGER, S. (2000) Unified maximum likelihood estimates for closed capture-recapture models using mixtures, *Biometrics*, **56**, pp. 443-450.
- POLLOCK, K. H. (1982) A capture-recapture design robust to unequal probability capture, *Journal of Wildlife Management*, **46**, 752-757.
- POLLOCK, K. H. & OTTO, M. C. (1983) Robust estimation of population size in closed animal populations from capture-recapture experiments, *Biometrics*, **39**, pp. 1035-1049.
- POLLOCK, K. H., HINES, J. E. & NICHOLS, J. D. (1984) The use of auxiliary variables in capture-recapture and removal experiments, *Biometrics*, **40**, pp. 329-340.
- POLLOCK, K. H., BUNCK, C. M. & CURTIS, P. D. (1989) Survival analysis in telemetry studies: the staggered entry design, *Journal of Wildlife Management*, **53**, pp. 7-15.
- POLLOCK, K. H., NICHOLS, J. D., BROWNIE, C. & HINES, J. E. (1990) Statistical inference for capture-recapture experiments, *Wildlife Society Monographs*, **107**, pp. 1-97.
- POLLOCK, K. H. (1991) Modelling capture, recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: past, present, and future, *Journal of the American Statistical Association*, **86**, pp. 225-238.
- POLLOCK, K. H., BUNCK, C. M., WINTERSTEIN, S. R. & CHEN, C. L. (1995) A capture-recapture survival analysis model for radio-tagged animals, *Journal of Applied Statistics*, **22**, pp. 661-672.
- POLLOCK, K. H. (2000) Capture-recapture models, *Journal of the American Statistical Association*, **95**, pp. 293-296.
- PRADEL, R. (1996) Utilization of capture-mark-recapture for the study of recruitment and population growth rate, *Biometrics*, **52**, pp. 703-709.
- RIGGS, M. R. & POLLOCK, K. H. (1992) A risk ratio approach to multivariable analysis of survival in longitudinal studies of wildlife populations. In: D. R. MCCULLOUGH & R. H. BARRETT (Eds), *Wildlife 2001: Populations*, pp. 74-89 (New York, Elsevier Applied Science).
- SANATHANAN, L. (1972). Estimating the size of a multinomial population, *The Annals of Mathematical Statistics*, **43**, pp. 142-152.
- SCHWARZ, C. J., SCHWEIGERT, J. F. & ARNASON, A. N. (1993) Estimating migration rates using tag-recovery data, *Biometrics*, **49**, pp. 177-193.
- SCHWARZ, C. J. & ARNASON, A. N. (1996) A general methodology for the analysis of capture-recapture experiments in open populations, *Biometrics*, **52**, pp. 860-873.
- SCHWARZ, C. J. & SEBER, G. A. F. (1999) A review of estimating animal abundance III, *Statistical Science*, **14**, pp. 427-456.
- SEBER, G. A. F. (1965) A note on the multiple recapture census, *Biometrika*, **52**, pp. 249-259.
- SEBER, G. A. F. (1982) *The Estimation of Animal Abundance and Related Parameters*, 2nd edn (London, C. W. Griffin).
- SMITH, S. G., SKALSKI, J. M., SCHLECHTE, J. W., HOFFMAN, A. & CASSEN, V. (1994) SURPH. 1

- Statistical survival analysis of fish and wildlife tagging studies. Center for Quantitative Sciences, University of Washington, Seattle.
- SPIEGELHALTER, D. J., THOMAS, A. & BEST, N. G. (1999) WinBUGS Version 1.2 User Manual MRC Biostatistics Unit, Cambridge. web site: <http://www.mrc-bsu.cam.ac.uk/bugs>
- SPENDELOW, J. A., NICHOLS, J. D., NISBET, I. C. T., HAYS, H., CORMONS, G. D., BURGER, J., SAFINA, C., HINES, J. E. & GOCHFELD, M. (1995) Estimating adult survival and movement rates of adults within a metapopulation of roseate terns, *Ecology*, **76**, pp. 2415-2428.
- STANLEY, T. R. & BURNHAM, K. P. (1998) Estimator selection for closed capture-recapture models, *Journal of Agricultural, Biological, and Ecological Statistics*, **3**, pp. 131-150.
- TRENT, T. T. & RONGSTAD, O. J. (1974) Home range and survival of cottontail rabbits in Southwestern Wisconsin, *Journal of Wildlife Management*, **38**, pp. 459-472.
- VAN DUESEN, P. C. (2001) An EM algorithm for capture-recapture experiments. *Journal of Environmental and Ecological Statistics*, (To appear).
- VOUNASTOU, P. & SMITH, A. F. M. (1995) Bayesian analysis of ring-recovery data via Markov Chain Monte Carlo simulation, *Biometrics*, **51**, pp. 687-728.
- WHITE, G. C. (1983) Numerical estimation of survival rates from band recovery and biotelemetry data, *Journal of Wildlife Management*, **47**, pp. 716-728.
- WHITE, G. C. & BURNHAM, K. P. (1999) Program MARK: survival estimation from populations of marked animals, *Bird Study*, **46**, pp. S120-S139
- WHITE, G. C., BURNHAM, K. P. & ANDERSON, D. R. (2000) Advanced features of Program MARK. Hungarian Symposium paper. MARK webpage reference.
- WHITE, G. C. (2001) Discussion. The use of auxiliary variables in capture-recapture modelling: an overview, In: B. J. T. MORGAN & D.L. THOMSON (Eds) *Statistical Analysis of Data From Marked Bird Populations*. *Journal of Applied Statistics* (This volume).
- YIP, P. S. F. (1989) An inference procedure for a capture and recapture experiment with time-dependent capture probabilities, *Biometrics*, **45**, pp. 471-479.
- YIP, P. S. F. (1991) A martingale estimating equation for a capture-recapture experiment in discrete time, *Biometrics*, **47**, pp. 1081-1088.
- YIP, P. S. F. & FONG, D. Y. T. (1993) Estimating population size by recapture sampling via estimating function. *Communications in Statistics—Stochastic Models*, **9**, pp. 179-193.
- YIP, P. S. F., WAN, E. C. Y. & CHAN, K. S. (2001) Robust design and use of auxiliary variables in estimating population size for an open population, *Biometrics* (To appear).