**7.5**

# ADVANCED FEATURES OF PROGRAM MARK

GARY C. WHITE

Department of Fishery and Wildlife Biology, Colorado State University, Fort Collins, CO 80523, USA, gwhite@cnr.colostate.edu

KENNETH P. BURNHAM

Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins, CO 80523, USA

DAVID R. ANDERSON

Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins, CO 80523, USA

*Abstract:* Program MARK provides a number of sophisticated analyses beyond just providing parameter estimates and their estimated conditional measures of precision. Advanced features include graphs of parameter estimates from 1 or more models, models constructed with the design matrix to estimate the mean of a set of real parameters, quasi-likelihood procedures to correct for overdispersion of the data, model averaging, variance components procedures to separate sampling and process variance in a set of parameter estimates, and use of the bootstrap procedure to evaluate goodness of fit.

*Key words:* AIC, Akaike weights, band recoveries, bootstrap, MARK, mark–recapture, model averaging, process variance, ringing recoveries, survival estimation

Monitoring biological populations is receiving increased emphasis even in less-developed areas of the world (Likens 1989). Estimation of survival probabilities, how they vary by age, sex, and time, and how survival might be correlated with external variables are difficult subjects. Estimation of immigration and emigration rates, population size, and the proportion of age classes that first enter the breeding population are equally important and difficult to estimate with precision for free-ranging populations. Estimation of the finite rate of population change ($\lambda$) and fitness ($F$) are still more difficult to address in a rigorous manner.

Risk assessment in higher vertebrates can be done in the framework of capture–recapture theory (Anderson et al. 1995). Population viability analyses must rely on estimates of vital rates of a population (Boyce 1992, White 2000); often these can be derived only from the study of marked animals. The richness component of biodiversity can often be estimated in the context of closed model capture–recapture (Nichols et al. 1998, Boulinier et al. 1998). Finally, the monitor-

ing components of adaptive management (Williams et al. 1996) can be rigorously addressed by data analysis from marked subpopulations.

Capture–recapture surveys have been used as a general sampling and analysis method to assess population status and trends in many vertebrate populations. The use of marked individuals is analogous to the use of various tracers in studies of physiology, medicine, and nutrient cycling. Recent advances in technology allow a wide variety of marking methods (e.g., Parker et al. 1990). Problems involving parameter estimation for threatened and endangered species and indicator species are common (e.g., Burnham et al. 1996). Banding and recovery methods have been used in many more-developed countries; these methods have many similarities with the capture–recapture surveys. Seber (1982, 1986, 1992) and Seber and Schwarz (1999) provide references on approximately 2,000 research papers on modeling and parameter estimation based on general capture methodology.

Program MARK (White and Burnham 1999) unifies these techniques and provides parameter estimates and associated standard errors for 17 classes of the models developed from data on the encounter histories of marked animals. MARK is based on likelihood theory (Edwards 1992), and all estimates are maximum likelihood estimates (MLEs). Models currently included in MARK are the live-recapture or Cormack-Jolly-Seber model (Cormack 1964, Jolly 1965, Seber 1965, Lebreton et al. 1992), dead recoveries or band recoveries with the Brownie et al. (1985) and Seber's (1970) parameterizations, joint live and dead encounters with Burnham's (1993) and Barker's (1997) parameterizations, known fate or radiocollar models (White and Garrott 1990), closed captures (Otis et al. 1978, White et al. 1982), robust design (Kendall and Nichols 1995; Kendall et al. 1995, 1997), multi-strata (Hestbeck et al. 1991, Brownie et al. 1993), and 3 parameterizations of Pradel's (1996) models. Input data to MARK are encounter histories that can consist of data from live recaptures, live resightings, or dead recoveries of the marked animals.

Here, we discuss the advanced features of MARK that allow sophisticated analyses of the parameter estimates. These features include graphs of parameter estimates from 1 or more models, models constructed with the design matrix to estimate the mean of a set of real parameters, quasi-likelihood procedures to correct for overdispersion of the data, model averaging to add model selection uncertainty into estimates of precision, variance components procedures to separate sampling and process variance in a set of parameter estimates, and use of the bootstrap procedure to evaluate goodness of fit.

## GRAPHICAL ANALYSIS

Once parameter estimates have been obtained for 1 or more models, estimates and their 95% confidence

intervals can be displayed graphically. The nested menu choices OUTPUT > SPECIFIC MODEL OUTPUT > INTERACTIVE GRAPHICS GS, or the Graphics Icon, will present a list of parameters for the model highlighted in the Results Browser window. The user enters indices of parameters to graph in an edit window; i.e., "1 to 9" or "15 to 17, 19 to 25," or clicks on parameters in the parameter list to select them to graph. Other options are available, such as combining the "to" operator with the "by" operator to select steps within the range. For example, 1 to 10 by 3, would select the values 1, 4, 7, and 10. Once parameters are selected, the OK button is clicked to generate the graph.

One option available once the graph is displayed is to add additional series of parameters to the graph. For example, the user might want to graph survival estimates and then add the reporting rates for the same years for estimates of encounter histories from dead recoveries. This graph would provide a display that might show a negative correlation between the estimates across time, suggesting that harvest of the species is causing additive mortality if the sampling covariance between estimates is accounted for by the analysis. A more likely scenario is that the user would want to graph survival estimates from 2 different models to compare their values. To specify the parameter estimates for the second model, the user must first highlight the desired model in the Results Browser (which will be different than the first model that was used to start the graph), and then click the Add Series button. The dialog window to specify parameters is once again shown, allowing the user to select which parameter values to graph.

The graphics display is produced by Graph Control™ published by Pinnacle Publishing Inc. This ActiveX control and its associated help file are installed in the \Windows\System subdirectory when MARK is installed. The icons displayed at the top of the graph window lead to a tabbed dialog window that allows extensive manipulation of the graph, including changing the data or format of the graph, copying the graph to the Windows clipboard, saving the graph and data in a format for later retrieval by Graph Control, and printing the graph. The features of Graph Control are extensive and allow the user complete control of the format of the graphics display (Fig. 1). An extensive help file is included with Graph Control to provide the user with the information needed to use the software.

## ESTIMATION OF MEAN VALUES WITH DESIGN MATRIX CODING

Often the parameter of interest for a series of survival estimates is an estimate of the mean survival rate. If the model with constant survival (i.e., $S$) fits the observed data much better than the time-specific model ($S_t$), then the constant survival estimate can be used. But, suppose that time-specific variation in the data is significant. In this section, we will demonstrate
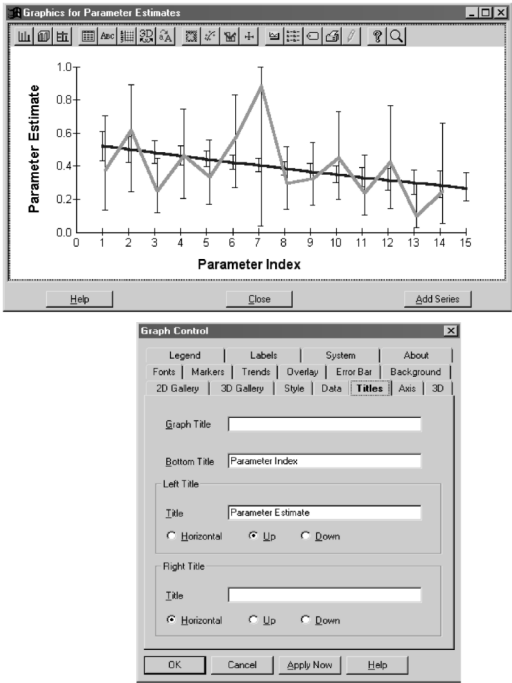


Fig. 1. Example of the graphics interface available in program MARK. The upper pane shows the plotted survival estimates for 2 models $S_t$ and $S_T$, whereas the lower pane provides the tools to manipulate the graph.

how to code the design matrix to obtain the mean of survival rates with time-specific variation.

To illustrate the technique, assume a design matrix with 5 estimates of survival rate. The default in MARK for a time-specific estimate of survival would be a $5 \times 5$ identity matrix. Another alternative is an intercept and 4 time effects:

|             | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ |
|-------------|-----------|-----------|-----------|-----------|-----------|
| $\hat{S}_1$ | 1         | 1         | 0         | 0         | 0         |
| $\hat{S}_2$ | 1         | 0         | 1         | 0         | 0         |
| $\hat{S}_3$ | 1         | 0         | 0         | 1         | 0         |
| $\hat{S}_4$ | 1         | 0         | 0         | 0         | 1         |
| $\hat{S}_5$ | 1         | 0         | 0         | 0         | 0         |

In this design matrix, the intercept ($\beta_1$) is the estimate of survival for the fifth occasion after the inverse link function is applied, i.e., $\hat{S}_5$. The survival estimate for the first occasion consists of the intercept and the offset provided by $\beta_2$, again after transforming by the inverse link function. Similarly $\beta_1 + \beta_3$ provides the survival estimate for $S_2$. By modifying the coding of the above design matrix, the intercept after back

transformation can become the mean of the survival estimates:

|        | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $\beta_5$ |
|--------|-----------|-----------|-----------|-----------|-----------|
| $\hat{S}_1$ | 1 | 1  | 0  | 0  | 0  |
| $\hat{S}_2$ | 1 | 0  | 1  | 0  | 0  |
| $\hat{S}_3$ | 1 | 0  | 0  | 1  | 0  |
| $\hat{S}_4$ | 1 | 0  | 0  | 0  | 1  |
| $\hat{S}_5$ | 1 | -1 | -1 | -1 | -1 |

With this design matrix, the back-transformed $\beta_1$ provides the mean of the survival estimates. The estimates of $\beta_2$ through $\beta_5$ provide the time variation around the estimated mean. To see this result, compute the mean of the rows of the matrix; i.e., $[(\beta_1 + \beta_2) + (\beta_1 + \beta_3) + (\beta_1 + \beta_4) + (\beta_1 + \beta_5) + (\beta_1 - \beta_2 - \beta_3 - \beta_4 - \beta_5)]/5 = 5\beta_1/5 = \beta_1$.

To obtain the estimate of mean survival, the estimate of $\beta_1$ must be converted with the link function used for the model. For example, with the logit link, $\bar{S} = 1/(1 + e^{-\beta_1})$ with standard error $SE(\bar{S}) =$

$$\left( \frac{\hat{\text{var}}(\hat{\beta}_1)\, e^{2\hat{\beta}_1}}{\left(1 + e^{\hat{\beta}_1}\right)^4} \right)^{1/2}$$

This SE provides the sampling variation of the estimate of the mean and does not include the process variance associated with the set of estimates. That is, this SE represents a fixed effects design, with the SE providing the precision of the mean for the observed time period. An alternate estimator of mean survival (discussed below) is developed as part of the variance components analysis, with the SE of the mean from that estimator including both sampling and process variance. Thus, the variance components estimator represents a random effects design, and hence is always larger than the SE provided by the approach presented here.

A nasty theoretical issue exists with this procedure—bias of the estimate of the mean as a result of transforming and back-transforming from the nonlinear link function. Consider an example: $S_i$ are 0.5, 0.6, 0.7, 0.8, and 0.9, so that $\bar{S} = 0.7$. With the logit link, the transformed values are 0, 0.40547, 0.8473, 1.38629, and 2.19722, giving an intercept of $\beta_1 = 0.96726$. Back transforming this value gives 0.7246, not 0.70. Thus, all the link functions in MARK except the identity link will provide slightly biased estimates of the mean value. Because the identity link is a linear function, estimates of the mean from the identity link will be unbiased. Typically, standard errors of the estimates will dominate such transformation bias so that the bias is usually ignored.

## QUASI-LIKELIHOOD ESTIMATION

The models considered in program MARK are based on products of multinomial likelihoods. An important,

implicit assumption of this statistical model is that events are independent; i.e., death of 1 animal does not depend on the fate of another. Another assumption is that each marked animal has the same survival probability. If these assumptions are violated, then a condition called "overdispersion" results. The survival rate is estimated without bias, but the variance estimate from the sample is too small. The observed data are more "dispersed" than is expected under the model. Quasi-likelihood estimation (Wedderburn 1974) provides a procedure to correct the estimates of sampling variance of the parameters and to allow for overdispersion in model selection.

A variance inflation factor, $c$, is estimated to correct the estimates of sampling variances and covariances. The parameter $c$ can be estimated from a goodness-of-fit chi-square statistic ($\chi^2$) and its degrees of freedom as $\hat{c} = \chi^2/df$ (Cox and Snell 1989, Lebreton et al. 1992, Tjur 1998). In MARK, the deviance of a model is asymptotically $\chi^2$ and provides a goodness-of-fit statistic that can sometimes be useful for evaluating goodness-of-fit and estimating $c$. Deviance is defined as –2 times the difference of the log likelihood of the model of interest and the saturated model for the data (McCullagh and Nelder 1989), where the maximum possible log likelihood is associated with the saturated model. Each degree of freedom of the data in a saturated model has an associated parameter, so that the saturated model is a perfect fit to the data. The deviance/df ratio is an estimate of $c$ for the dead recovery models. Unfortunately, for live recapture models, deviance/df is a poor estimate of $c$, mainly because there are so many potential encounter histories that can be realized, but generally only a few are actually observed and provided in the encounter histories file. As a result, the deviance is not well approximated by the asymtotic distribution. In the section on bootstrap procedure below, we will consider how to obtain reasonable estimates of $c$. For now, assume that some estimate of $c$ is available.

When $c$ is equal to 1, MARK uses $AIC_c$ for model selection (Burnham and Anderson 1998),

$$AIC_c = -2\log(\mathcal{L}(\hat{\beta}) + 2K + \frac{2K(K+1)}{n-K-1},$$

where $\log(\mathcal{L}(\hat{\beta}))$ is the log of the likelihood of the parameters ($\beta$) given the data, $K$ is the number of parameters estimated, and $n$ is the finite sample size (see Burnham and Anderson 1998 for additional discussion of $AIC_c$). When overdispersion exists in the data, i.e., $c > 1$, quasi-likelihood theory suggests the modification (Burnham and Anderson 1998) as:

$$QAIC_c = \frac{-2\log(\mathcal{L}(\hat{\beta}))}{\hat{c}} + 2K + \frac{2K(K+1)}{n-K-1}.$$

In MARK, the value of $\hat{c}$ is set with the ADJUSTMENTS > C-HAT menu choices from the Results Browser window.

You can tell when $\hat{c}$ is >1 because the headings of the columns in the Results Browser will be QAIC$_c$ instead of AIC$_c$, and the model ordering will reflect the value of $\hat{c}$.

The value of $\hat{c}$ also will be used to inflate the variances and covariances of parameter estimates. When either real or beta parameter estimates are listed with the OUTPUT > SPECIFIC MODEL OUTPUT > REAL ESTIMATES or OUTPUT > SPECIFIC MODEL OUTPUT > BETA ESTIMATES from the Results Browser window (or their associated icons), the output listing will include the value of $\hat{c}$ in headings, and the standard errors reported will have been inflated by multiplying by $\sqrt{\hat{c}}$. Note that the correction for $\hat{c}$ is applied to any further use of the parameter estimates and their associated standard errors, such as model averaging, variance components estimation, or graphics. However, the standard errors in the full output for the model produced when the parameters were originally estimated will not reflect the value of $\hat{c}$ because the numerical estimation output has no knowledge of the value of $\hat{c}$. Thus, if you set $\hat{c} > 1$, the standard errors in the output window obtained by the menu choice OUTPUT > SPECIFIC MODEL OUTPUT > OUTPUT IN NOTEPAD will differ from the standard errors reported by OUTPUT > SPECIFIC MODEL OUTPUT > REAL ESTIMATES or OUTPUT > SPECIFIC MODEL OUTPUT > BETA ESTIMATES. Only the latter have been inflated by $\sqrt{\hat{c}}$.

Only 1 value of $\hat{c}$ is used for all models in a set of models. This value should be estimated from the global or most general model and applied to the remainder of the models in the set of models considered in the analysis.

## MODEL AVERAGING

Usually, several models seem plausible, based on the AIC$_c$ or QAIC$_c$ values. In this case, a formal way bases inference on more than a single model. Model averaging, available by the OUTPUT > MODEL AVERAGING menu choices from the Results Browser window, allows you to compute the average of a real parameter from the models in the Results Browser. By doing so, you include model selection uncertainty in the estimate of precision of the parameter, and thus produce unconditional estimates of sampling variances and covariances and standard errors. Parameters that are to be model-averaged are selected from the possible parameters available in the parameter index matrices (PIMs). Thus, only real parameters can be selected because the beta parameters do not occur in the PIMs, and a given $\beta_i$ does not occur in all models.

Model averaging entails a weighted average of the estimates of a parameter for $R$ models shown in the Results Browser window. Akaike weights, $w_i$, are a natural weight to use. The Akaike weight for model (Burnham and Anderson 1998) is defined as:

$$w_i = \frac{\exp(-\Delta_i/2)}{\sum\limits_{j=1}^{R} \exp(-\Delta_j/2)}$$

where $\Delta_i$ is the difference in AIC$_c$ value for model $i$ and the minimum AIC$_c$ model. Both $\Delta_i$ and $w_i$ are columns in the MARK Results Browser. If the Akaike weights do not appear in the Results Browser, then change the FILE > PREFERENCES menu choice to show the $w_i$ the next time the Results Browser is started. Alternative weights can come from estimates of model selection frequencies (e.g., weights based on the bootstrap technique; Burnham and Anderson 1998), but this option currently is not available in MARK.

The precision of an estimator ideally should have 2 variance components: (1) the conditional sampling variance, given a model $M_i$ with sampling variance $\hat{\text{var}}(\hat{\theta}|M_i)$ and (2) variation associated with model selection uncertainty. Buckland et al. (1997) provide an effective method to estimate precision not conditional on a particular model. Assume that the scalar parameter of interest is a real parameter, and hence, must be common to all models considered (e.g., $\phi$, or $N$, or $S$). If our focus is on a structural parameter of the model (i.e., a beta parameter in MARK) that appears only in a subset of our full set of models, then we must restrict ourselves to that subset to make the sort of inferences considered here about the parameter of interest. MARK will automatically model-average real parameters, but the user must do the calculations manually to model average beta parameters.

From Buckland et al. (1997) we will take the estimated unconditional $\hat{\text{var}}(\bar{\bar{\theta}})$ as:

$$\hat{\text{var}}\left(\bar{\bar{\theta}}\right) = \left( \sum_{i=1}^{R} w_i \sqrt{\hat{\text{var}}(\hat{\theta}_i|M_i) + \left(\hat{\theta}_i - \bar{\bar{\theta}}\right)^2} \right)^2$$

where $\bar{\bar{\theta}} = \sum\limits_{i=1}^{R} w_i \hat{\theta}_i$, and the $w_i$ are the Akaike weights scaled to sum to 1 as defined above. The subscript $i$ refers to the $i$th model. $\bar{\bar{\theta}}$ is a weighted average of the estimated parameter over $R$ models ($i = 1, 2, ..., R$). This estimator of the unconditional variance is clearly the sum of 2 components: the conditional sampling variance $\hat{\text{var}}(\hat{\theta}|M_i)$ and a term for the variation in the estimates across the $R$ models, $(\hat{\theta}_i - \bar{\bar{\theta}})^2$. The square root of the sum of these terms is then merely weighted by the $w_i$. The estimated unconditional $\hat{\text{SE}}(\bar{\bar{\theta}}) = \sqrt{\hat{\text{var}}(\bar{\bar{\theta}})}$. (Burnham and Anderson (1998) found this estimator to be robust to imprecise values of the weights, and suggest it as appropriate for scenarios such as found in program MARK.

A logical and simple $(1 - \alpha)100\%$ unconditional confidence interval when sample size is large is given by the end points $\bar{\bar{\theta}} \pm z_{1-\alpha/2}\hat{\text{SE}}(\bar{\bar{\theta}})$, and is the default confidence interval provided by MARK (Fig. 2). Modifications were suggested by Burnham and Anderson (1998). We suggest incorporating additional information into the confidence interval when possible, such as the minimum number known alive ($M_{t+1}$) for estimates of population size ($N$). For example, the log-based confidence interval used in program CAPTURE for $\hat{N}$ is:

$$[M_{t+1} + (\hat{N} - M_{t+1}) / C, \; M_{t+1} + \hat{N} - M_{t+1}) \times C],$$

where

$$C = \exp\!\left( z_{1-\alpha/2} \sqrt{ \log\!\left[ 1 + \frac{\hat{\text{var}}(\hat{N})}{(\hat{N} - M_{t+1})^2} \right] } \right).$$

This procedure requires $\hat{N} > M_{t+1}$.

Another example is using the logistic transformation for survival estimates so that the confidence interval is within the feasible interval of 0 to 1. The variance of the $\text{logit}(\bar{\hat{\theta}})$ is

$$\frac{\hat{\text{var}}(\bar{\hat{\theta}}),}{\bar{\hat{\theta}}^2 (1 - \bar{\hat{\theta}})^2},$$

so that confidence intervals can be computed on $\text{logit}(\bar{\hat{\theta}})$ and then back-transformed resulting in a bounded confidence interval (Burnham et al. 1987:214), i.e.,

$$\frac{1}{1 + \exp\!\left[ -\left( \text{logit}(\bar{\hat{\theta}}) - z_{1-\alpha/2} \sqrt{\hat{\text{var}}[\text{logit}(\bar{\hat{\theta}})]} \right) \right]},$$

$$\frac{1}{1 + \exp\!\left[ -\left( \text{logit}(\bar{\hat{\theta}}) + z_{1-\alpha/2} \sqrt{\hat{\text{var}}[\text{logit}(\bar{\hat{\theta}})]} \right) \right]},$$

A different form of the same confidence interval that may provide more insight into the equation is given by:

$$C = \exp\!\left[ z_{1-\alpha/2} \sqrt{\hat{\text{var}}[\text{logit}(\bar{\hat{\theta}})]} \right], \text{ so that}$$

$$\hat{\theta}_L = \frac{\hat{\theta}}{\bar{\hat{\theta}} + (1 - \bar{\hat{\theta}})/C} \; \text{ and } \; \hat{\theta}_U = \frac{\hat{\theta}}{\bar{\hat{\theta}} + (1 - \bar{\hat{\theta}})C}.$$

## VARIANCE COMPONENTS

Program MARK can compute an estimate of the underlying process variance, $\sigma^2$, for a set of parameter estimates. Either the real parameters or the beta parameters may be used. Select either VARIANCE COMPONENTS REAL PARAMETERS or VARIANCE COMPONENTS BETA PARAMETERS from the OUTPUT > SPECIFIC MODEL OUTPUT menu choice of the Results Browser window. To describe the method, we will assume that we are estimating the process variance of a set of survival probabilities. The basic concept is that the $S_1, S_2, ..., S_n$, are considered a random sample from some distribution, hence, have a mean ($\mu$) and variance ($\sigma^2$). That is, the values are assumed uncorrelated. If we could directly observe (i.e., measure without error) these survival rates, then we would make use of the $S_i$ in our models on population dynamics. We would also compute a mean

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} S_i \text{ and the usual estimate of}$$



Fig. 2. Example of model averaging output from program MARK. The column labeled "Standard Error" is the estimate of sampling standard error for the model. The row labeled "Weighted Average" is the model averaging estimate of the parameter value, and the next row provides the unconditional SE of this estimate. In this example, $13.71\% \; (= 100 \times (1 - 0.0615626^2/0.0662732^2)$ of the variance of the model averaging estimate is attributable to model uncertainty. If inferences were made only from the estimate from the minimum $\text{AIC}_c$ model, the standard error of the estimate would have been 0.0463568, compared to the unconditional standard error of 0.0662732, an inflation of 43%.

$\hat{\sigma}^2 = \dfrac{1}{n-1} \sum\limits_{i=1}^{n} \left( S_i - \bar{S} \right)^2.$ However, we have instead a type of measurement error variation and covariation in our estimated values $\hat{S}_i$. When that measurement (sampling) variation is included into the inference methods, we have more complicated estimators of these 2 "population" parameters. Moreover, if the estimated sampling variances and covariances in the real parameters variance–covariance matrix ($W$) are too big then we cannot reliably partition the total observed variance,

$$\frac{1}{n-1} \sum_{i=1}^{n} \left( S_i - \bar{S} \right)^2,$$

into $\hat{\sigma}^2$ and the average of the diagonal elements minus the average of the nondiagonal elements of the real parameters variance–covariance matrix.

The method used in program MARK assumes that the vector of parameter estimates ($\hat{S}$) was generated from a general model (i.e., the model did not put constraints on the parameter estimates that would affect the process variance estimate). First, we assume $\hat{S} = S + \delta$, given $S$. Conditional on $S$, $\delta$ has a variance–covariance matrix $W$, hence, $E(\hat{S}|\underline{S}) = \underline{S}$ for large samples. Second, unconditionally $S$ is a random vector with expectation $X\beta$ and variance–covariance matrix $\sigma^2 I$, where $I$ is the identity matrix. (The vector $\beta$ is different than the beta parameters of the MARK link function). Thus, the process errors $\varepsilon_i = S_i - E(S_i)$ are independent with homogeneous variance $\sigma^2$. Also, we assume mutual independence of sampling errors $\delta$ and process errors $\varepsilon$. We fit a model that does not constrain $S$, e.g., $\{S_t\}$, and hence get the maximum likelihood

estimates $\hat{S}$ and an estimate of W. Let $S$ be a vector with $n$ elements, and $\beta$ have $k$ elements.

Unconditionally,

$$\hat{S} = X\beta + \delta + \varepsilon, \; VC(\delta + \varepsilon) = D = \sigma^2 I + W.$$

We want to estimate $\beta$ and $\sigma^2$, an unconditional variance–covariance matrix for $\hat{\beta}$, a confidence interval on $\sigma^2$, and to compute a shrinkage (i.e., empirical Bayes or random effects) estimator of $S$ ($\tilde{S}$) and its conditional sampling variance–covariance matrix. In this random effects context the maximum likelihood estimator is the best conditional estimator of $S$. However, once we add the random effects structure we can consider an unconditional estimator of $S$ ($\tilde{S}$) and a corresponding unconditional variance–covariance for $\tilde{S}$, which incorporates $\sigma^2$ as well as $W$ and has $n - k$ degrees of freedom (if we are assuming large df for $W$ and "large" $\sigma^2$).

For a given $\sigma^2$ we have

$$\hat{\beta} = (X'D^{-1}X)^{-1}X'D^{-1}\hat{S}.$$

Note that here $D$, hence $\hat{\beta}$, is a function of $\sigma^2$. Now we need a criterion that allows us to find an estimator of $\sigma^2$. By the method of moments we have

$$n - k = (\hat{S} - X\hat{\beta})^{-1}X'D^{-1}(\hat{S} - X\hat{\beta}). \quad (1)$$

The above equation defines a 1-dimensional numerical-solution search problem. Pick a value of $\sigma^2$, compute $D$, then compute $\hat{\beta}$, then compute the right-hand side of the equation (1); this process is repeated over values of $\sigma^2$ until the solution of equation (1), as $\tilde{S}$, is found. This process also gives $\hat{\beta}$. The process can be simplified by eliminating $\hat{\beta}$ from equation (1). Define

$$A = X(X'D^{-1}X)^{-1}X'.$$

This is the "hat" matrix of linear regression, but we choose to not call it $H$. Now equation (1) is

$$n - k = -\hat{S}'D^{-1} - D^{-1}AD^{-1})\hat{S}.$$

The unconditional variance–covariance matrix of $\hat{\beta}$ is

$$VC(\hat{\beta}) = (X'D^{-1}X)^{-1}.$$

A distributional assumption is needed to get a confidence interval on $\sigma^2$, although a bootstrap procedure could be used. One approach is to assume a pivotal

$$SRSS(\sigma^2) = (\hat{S} - X\hat{\beta})'D^{-1}(\hat{S} - X\hat{\beta})$$

is a central $\chi^2$ on df $= n - k$. A 95% confidence interval on $\sigma^2$ is found as the solution points (lower and upper bounds on $\sigma^2$, respectively) of SRSS = upper 97.5%-tile point of a central and SRSS = lower 2.5%-tile point of a central $\chi^2_{n-k}$. The lower bound should be

taken as zero if the solution would otherwise be negative. We expect the upper bound to always be finite.

Define another matrix as

$$H = \sigma D^{-1/2} = \sigma(\sigma^2 I + W)^{-1/2} = (I + \frac{1}{\sigma^2}W)^{-1/2};$$

we only need $H$ at $\hat{\sigma}$. The shrinkage estimate of $S$ used in program MARK is

$$\tilde{S} = H(\hat{S} - X\hat{\beta}) + X\hat{\beta},$$
$$= H\hat{S} + (I - H)X\hat{\beta}.$$

To get an estimator of the variance of these shrinkage estimators (not exact as the estimation of $\sigma^2$ is ignored here, as it is for the variance–covariance matrix of $\hat{\beta}$), we can write

$$\tilde{S} = [H + (I - H)AD^{-1}]\hat{S} = G\hat{S}.$$

The conditional variance–covariance matrix of the shrinkage estimator is then $VC(\tilde{S}|\underline{S}) = GWG'$, whereas $W = VC(\tilde{S}|\underline{S})$. Because $\tilde{S}$ is known to be biased and the direction of the bias is known, a much improved basis for inference is $VC(\tilde{S}|\hat{S}) = GWG' + (\tilde{S} - \hat{S})(\tilde{S} - \hat{S})'$. The square roots of the diagonal elements of this matrix are

$\hat{rmse}(\tilde{S}_i|\underline{S}) = \sqrt{\hat{var}(\tilde{S}_i|\underline{S}) + (\tilde{S}_i - \hat{S}_i)^2}$. Confidence intervals on $S_i$ should be based on this $\hat{rmse}(\tilde{S}_i|\underline{S})$, which is what MARK does. The term $\hat{rmse}(\tilde{S}_i|S_i)$ can exceed $SE(\hat{S}_i|S_i)$, but on average $\hat{rmse}(\tilde{S}_i|\underline{S})$ is smaller.

This set of calculations has been fully programmed in MARK. To specify the estimates ($\hat{S}_i$) to be used in estimating $\hat{\sigma}^2$, the same type of dialog window as used for selecting estimates to graph is used. See the section on selecting graph parameters above for details.

The default model for the $\hat{S}_i$ values is the mean (i.e., $X$ consists of a vector of 1's). To select a linear trend model, click that model in the upper right corner of the dialog window. You also can specify your own design matrix for $X$, instead of accepting 1 of the 2 predefined models.

The initial output from the variance component estimation will be displayed from the numerical procedure that computes the estimates. A graph of the original estimates and their confidence intervals, plus the "shrunk" estimates ($\tilde{S}$) will be displayed (Fig. 3). Tiled behind this graph is a listing of the "shrunk" estimates. After you close this set of windows, the full output from the variance component estimation routine will be displayed in a NotePad window, including the estimate of $\hat{\sigma}^2$ and its 95% confidence interval, $\hat{\sigma}$ and its 95% confidence interval, the vector $\hat{\beta}$ (the estimated parameters for the linear model) and its variance–covariance matrix, and the vector $\tilde{S}$ and its variance–covariance matrix and standard errors (Fig. 4). The column labeled RMSE(S-tilde) are the values that would normally be used for computing confidence intervals on $\tilde{S}$.
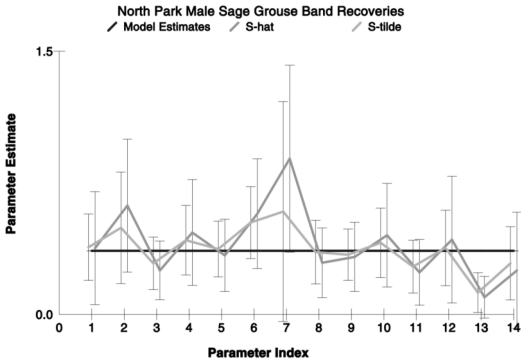
Fig. 3. Example of the variance components analysis from program MARK. The lighter line and confidence intervals are for $\hat{S}$, whereas the darker line and confidence intervals are for the shrinkage estimator $\tilde{S}$.

The variance reported for $\bar{S}$ from the variance components analysis can be written as:

$$\hat{\text{var}}(\bar{S}) \simeq \frac{\sigma^2}{n} + \frac{1}{n^2}\sum_{i=1}^{n}\hat{\text{var}}(\hat{S}_i|S_i)$$

where the variance for $\bar{S}$ provided by the design matrix approach discussed above would be just

$$\hat{\text{var}}(\bar{S}) \simeq \frac{1}{n^2}\sum_{i=1}^{n}\hat{\text{var}}(\hat{S}_i|S_i)$$

The procedure used in MARK differs from the procedures presented by Link and Nichols (1994) and Gould and Nichols (1998) in that the estimator in MARK weights the estimates of $\hat{S}$ by their estimated



Fig. 4. Portion of the numerical output for the variance components analysis from program MARK. Estimates of $\hat{\beta}$ and its unconditional variance–covariance (i.e., this variance–covariance matrix has $\hat{\sigma}^2$ included as well as sampling variation), $\hat{S}$ and associated sampling SE, $\tilde{S}$ and associated process SE, and both $\hat{\sigma}^2$ and $\hat{\sigma}$ with 95% confidence intervals. The estimates in the column labeled SE(S-tilde) are too small, so that the standard error of $\tilde{S}$ should be taken from the column labeled RMSE(S-tilde).

sampling variance–covariance matrix ($W$) whereas the other procedures give equal weight to each estimate. The procedure in MARK is an extension of the procedure described in Burnham et al. (1987).

## BOOTSTRAP GOODNESS-OF-FIT PROCEDURE

The goodness-of-fit of the global model can be evaluated in 3 ways: (1) assume that the deviance for the model is $\chi^2$ distributed and compute a goodness-of-fit test from this statistic; (2) use program RELEASE (for live-recapture data with no age effects and only time- and group-specific survival and recapture rates) to compute the goodness-of-fit tests provided by that program and developed in Burnham et al. (1987); or (3) use the parametric bootstrap procedure provided in MARK.

The first approach generally is not valid because the assumption of the deviance being approximately $\chi^2$ distributed seldom is met. This approach only seems to be reasonable for band recovery data sets with large numbers of animals marked. This approach has never been reasonable for live-recapture data because the number of potential encounter histories is so large that most of them are not observed, even for studies with a large number of animals marked.   Use of program RELEASE for live animal encounter studies is reasonable, but usually lacks statistical power to detect lack-of-fit because of the amount of pooling required to compute chi-square distributed test statistics.  Further, these tests use a restricted set of models. For example, RELEASE TEST3 asks whether the $m_{ij}$ array is sufficient for the data set concerned, and TEST2 is conditional on the $m_{ij}$ array being sufficient. Although goodness-of-fit procedures like that programmed in RELEASE could be developed theoretically for other data types, such as joint live and dead recoveries (Barker 1995), they generally have not been to date. For these reasons, the bootstrap procedure was implemented in MARK under the Results Browser Tests menu.   The bootstrap procedure provides a method to assess goodness-of-fit of all the data types in MARK, and provides $\hat{c}$ for inflating variances and quasi-likelihood model selection.

With the bootstrap procedure, the estimates of the model being evaluated for goodness-of-fit (generally something like a "global" model) are used to generate bootstrap data sets (i.e., a parametric bootstrap). The simulated data exactly meet the assumptions of the model (i.e., no overdispersion is included, animals are totally independent, and no violations of model assumptions are included). Simulated data are generated based on the number of animals released at each occasion. For each release, a simulated encounter history is constructed, conditional on first release. As an example, consider a live-recapture data set with 3 occasions (2 survival intervals) and an animal first released at time 1.   The animal starts off with an
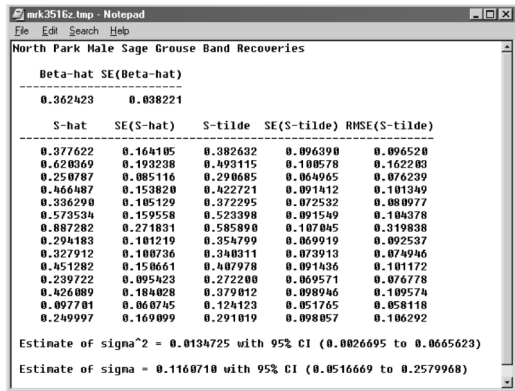
encounter history of 100 because it was released on occasion 1. Does the animal survive the interval from the release occasion until the next recapture occasion? The probability of survival is $\phi_1$, provided from the estimates obtained with the original data. A uniform random number in the interval (0, 1) is generated, and compared to $\hat{\phi}_1$ from the original data. If the random number is less than or equal to $\hat{\phi}_1$, then the animal is considered to have survived the interval. If the random value is greater than $\hat{\phi}_1$, then the animal has died. Thus, the encounter history would be complete and would be 100. Suppose instead that the animal survives the first interval. Then, is it recaptured on the second occasion? Again, a new random number is generated and compared to the capture probability $\hat{p}_2$ from the parameter estimates of the model being tested. If the random value is less than $\hat{p}_2$, then the animal is considered to be captured, and the encounter history would become 110. If not captured, the encounter history would remain 100. Next, whether the animal survives the second survival interval is determined, again by comparing a new random value with $\hat{\phi}_2$. If the animal dies, the current encounter history is complete and would be either 100 or 110. If the animal lives, then a new random value is used to determine if the animal is recaptured on occasion 3 with probability $\hat{p}_3$. If recaptured, the third occasion in the encounter history is given a 1. If not recaptured, the third occasion is left with a zero value.

Once the encounter history is complete, it is saved for input to the numerical estimation procedure. Once encounter histories have been generated for all the animals released, the numerical estimation procedure is run to compute the deviance and its degrees of freedom. These values along with $\hat{c}$ (= deviance / df) are saved to a simulation output file. The entire process is repeated for the number of simulations requested, often 1,000.

In MARK, the bootstrap goodness-of-fit procedure is accessed from the Results Browser with the menu choices TESTS > BOOTSTRAP GOF. Parameter estimates will be taken from the currently highlighted model in the Results Browser window, so you should click on the appropriate model before selecting the bootstrap procedure. You have a choice of only computing the deviance of the bootstrap simulations or of also computing $\hat{c}$. Computing $\hat{c}$ requires that the number of parameters estimated in the model be computed from the variance–covariance matrix of the parameter estimates, which can be quite expensive in terms of computer time. Thus, only taking the deviance from each simulation is the fastest approach. You next have to specify the name of a file to hold the simulation output. Typically, a good choice is a file name with the same name as the MARK results file you are operating from, but with GOF added to the end of the name to signify that the file contains goodness-of-fit results. Next, you specify the number of bootstrap simulations

to perform. The default is 100, but you may want to just run 1 or 2 to see if the procedure works correctly for your problem specification. When you click OK to accept the number of simulations shown in this dialog, MARK begins the simulation process. You will notice that the numerical estimation procedure is being executed for each simulation, and a progress bar shows the number of the simulations completed.

When the simulations are completed, the progress bar disappears. To view the results, you must select SIMULATIONS > VIEW SIMULATION RESULTS from the Results Browser window, and then select the file that you previously specified to hold the simulation results. A browser window will appear with the simulation results. You can manipulate the bootstrap results in several ways.

You will probably want to first sort the file by 1 of the variables, likely deviance. You do this by clicking on the button with the "A–Z down arrow" label, and selecting the variable to sort on as deviance. You can then locate the position in the file where the deviance from the original data would fall and identify the ranking of this value by clicking the icon with the diagonal arrow. This ranking can be used as the $P$ value of a goodness-of-fit test. For example, suppose that the deviance of the original model was 101.01, whereas the largest deviance from 1,000 simulations was only 90.90. Then you can conclude that the probability of observing a value as large as 101.01 was less than 1/1,000 under the null hypothesis. As another example, suppose the 801th simulated deviance in the sorted deviance file is 100.90, and the 802nd value was 101.50. Then, you would conclude that your observed deviance was reasonably likely to be observed, with probability of 198/1,000 (because 198 of the simulated values exceeded the observed value).

A similar procedure may be used to evaluate the observed $\hat{c}$ by comparing its rank to the simulated values of $\hat{c}$. Typically, conclusions using $\hat{c}$ and deviance are equivalent, but different results may be obtained with sparse data sets where the degrees of freedom associated with the deviance vary a lot across the simulations. We prefer to use the procedure based on deviance because this approach avoids the numerical difficulties of computing the number of parameters estimated.

The bootstrap simulations also can be used to estimate the overdispersion parameter, $c$. Two approaches are possible, either based on the deviance directly, or on $\hat{c}$. For the approach based on deviance, the deviance estimate from the original data is divided by the mean of the simulated deviances to compute $\hat{c}$ for the data. The logic is that the mean of the simulated deviances represents the expected value of the deviance under the null model of no violations of assumptions (i.e., perfect fit of the model to the data). Thus, $\hat{c}$ = observed deviance divided by expected deviance provides a measure of the amount of overdispersion in the original data. Typically, you would

examine the standard error of the mean deviance to be sure that the value is estimated adequately based on the number of simulations conducted. The third button in the simulation results window of immediate interest is the "Calculator" button, which will compute the mean, SE, and confidence interval for each of the variables in the file. Thus, you can compute the mean of the simulated deviances and compute the value of $\hat{c}$.

The second approach to estimating $\hat{c}$ for the original data is to divide the observed value of $\hat{c}$ from the original data by the mean of the simulated values of $\hat{c}$ from the bootstraps. Again, the mean of the simulated values provides an estimate of the expected value of $\hat{c}$ under the assumption of perfect fit of the model to the data.

We are not sure the benefits/disadvantages of the 2 procedures, but results are about the same for data sets with a reasonable number of animals. Results can be different when the degrees of freedom of the deviance varies a lot across the bootstrap simulations (caused by few releases). This discrepancy suggests that the deviance of the observed data divided by the mean of the simulated deviances is the preferred approach.

Currently, several limitations exist with the bootstrap procedure as implemented in program MARK. The procedure does not handle losses on capture (i.e., for animals with live recaptures where animals may be accidentally killed in the trapping process). Second, the observed values of individual covariates are not incorporated into the analysis. Third, the only way to estimate $\hat{c}$ for the Pradel (1996) models is to assume that the lack-of-fit is from the recaptures of previously marked animals. Thus, the data must be analyzed with the live-recaptures model and $\hat{c}$ estimated from that analysis.

Estimates of $c$ can be <1. We suggest not using $\hat{c}$ < 1, but rather truncating $\hat{c}$ to 1.

## FUTURE IMPROVEMENTS

Program MARK is being improved. Several enhancements are expected with respect to the features discussed here. Some alternative estimators of process variance likely will be implemented. Better graphics will be developed for the variance components procedure. Improvements in the bootstrap procedure to handle individual covariates and losses on capture will be developed. The simulation capability of MARK, of which the bootstrap procedure is a small part, will be improved. A capability for users to incorporate new data types and model structures into MARK will be developed, which will lead to many new models.

## CONCLUSION

An extensive and growing set of procedures are available in program MARK. These procedures provide sophisticated analyses of parameter estimates obtained from encounter history data. The analysis of marked animal encounter data combined with these advanced procedures will provide the answers needed for maintaining and managing numerous species in the future.

## LITERATURE CITED

ANDERSON, D. R., G. C. WHITE, AND K. P. BURNHAM. 1995. Some specialized risk assessment methodologies for vertebrate populations. Environmental and Ecological Statistics 2:91–115.

BARKER, R. J. 1995. Open population mark–recapture models including ancillary sightings. Dissertation, Massey University, Palmerston North, New Zealand.

———. 1997. Joint modeling of live-recapture, tag-resight, and tag-recovery data. Biometrics 53:666–677.

BOULINIER, T., J. D. NICHOLS, J. R. SAUER, J. E. HINES, AND K. H. POLLOCK. 1998. Estimating species richness to make inferences in community ecology: the importance of heterogeneity in species detectability. Ecology 79:1018–1028.

BOYCE, M. S. 1992. Population viability analysis. Annual Review Ecology Systematics 23:481–506.

BROWNIE, C., D. R. ANDERSON, K. P. BURNHAM, AND D. S. ROBSON. 1985. Statistical inference from band recovery data a handbook. Second edition. U.S. Fish and Wildlife Service, Resource Publication 156. Washington, D.C., USA.

———, J. E. HINES, J. D. NICHOLS, K. H. POLLOCK, AND J. B. HESTBECK. 1993. Capture–recapture studies for multiple strata including non-Markovian transitions. Biometrics 49:1173–1187.

BUCKLAND, S. T., K. P. BURNHAM, AND N. H. AUGUSTIN. 1997. Model selection: an integral part of inference. Biometrics 53:603–618.

BURNHAM, K. P. 1993. A theory for combined analysis of ring recovery and recapture data. Pages 199–213 in J.-D. Lebreton and P. M. North, editors. Marked individuals in the study of bird population. Birkhäuser Verlag, Basel, Switzerland.

———, AND D. R. ANDERSON. 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag, New York, New York, USA.

———, ———, AND G. C. WHITE. 1996. Meta-analysis of vital rates of the northern spotted owl. Studies in Avian Biology 17:92–101.

———, ———, ———, C. BROWNIE, AND K. H. POLLOCK. 1987. Design and analysis methods for fish survival experiments based on release–recapture. American Fisheries Society Monograph 5:1–437.

CORMACK, R. M. 1964. Estimates of survival from the sighting of marked animals. Biometrika 51:429–438.

COX, D. R., AND E. J. SNELL. 1989. Analysis of binary data. Second edition. Chapman & Hall, New York, New York, USA.

EDWARDS, A. W. F. 1992. Likelihood: expanded edition. Johns Hopkins University Press, Baltimore, Maryland, USA.

GOULD, W. R., AND J. D. NICHOLS. 1998. Estimation of temporal variability of survival in animal populations. Ecology 79:2531–2538.

HESTBECK, J. B., J. D. NICHOLS, AND R. A. MALECKI. 1991. Estimates of movement and site fidelity using mark-resight data of wintering canada geese. Ecology 72:523–533.

JOLLY, G. M. 1965. Explicit estimates from capture-recapture data with both death and immigration stochastic model. Biometrika 52:225–247.

KENDALL, W. L., AND J. D. NICHOLS. 1995. On the use of secondary capture–recapture samples to estimate temporary emigration and breeding proportions. Journal of Applied Statistics 22:751–762.

———, ———, AND J. E. HINES. 1997. Estimating temporary emigration using capture–recapture data with Pollock's robust design. Ecology 78:563–578.

———, K. H. POLLOCK, AND C. BROWNIE. 1995. A likelihood-based approach to capture–recapture estimation of demographic parameters under the robust design. Biometrics 51:293–308.

LEBRETON, J-D., K. P. BURNHAM, J. CLOBERT, AND D. R. ANDERSON. 1992. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. Ecological Monographs 62:67–118.

LIKENS, G. E., editor. 1989. Long-term studies in ecology. Springer-Verlag, Berlin, Germany.

LINK, W. A., AND J. D. NICHOLS. 1994. On the importance of sampling variance to investigations of temporal variation in animal population size. Oikos 69:539–544.

MCCULLAGH, P., AND J. A. NELDER. 1989. Generalized linear models. Second edition. Chapman & Hall, New York, New York, USA.

NICHOLS, J. D., T. BOULINIER, J. E. HINES, K. H. POLLOCK, AND J. R. SAUER. 1998. Inference methods for spatial variation in species richness and community composition when not all species are detected. Conservation Biology 12:1390–1398.

OTIS, D. L., K. P. BURNHAM, G. C. WHITE, AND D. R. ANDERSON. 1978. Statistical inference from capture data on closed animal populations. Wildlife Monographs 62.

PARKER, N. C., A. E. GIORGI, R. C. HEIDINGER, D. B. JESTER, JR., E. D. PRINCE, AND G. A. WINANS, editors. 1990. Fish-marking techniques. American Fisheries Society Symposium 7, Bethesda, Maryland, USA.

PRADEL, R. 1996. Utilization of capture–mark–recapture for the study of recruitment and population growth rate. Biometrics 52:703–709.

SEBER, G. A. F. 1965. A note on the multiple recapture census. Biometrika 52:249–259.

———. 1970. Estimating time-specific survival and reporting rates for adult birds from band returns. Biometrika 57:313–318.

———. 1982. The estimation of animal abundance and related parameters. Second edition. Macmillan, New York, New York, USA.

———. 1986. A review of estimating animal abundance. Biometrics 42:267–292.

———. 1992. A review of estimating animal abundance II. Reviews of the International Statistics Institute 60:129–166.

———, AND C. SCHWARZ. 1999. Review of estimating animal abundance III. Statistical Science 14:427–456.

TJUR, T. 1998. Nonlinear regression, quasi likelihood, and overdispersion in generalized linear models. The American Statistician 52:222–227.

WEDDERBURN, R. W. M. 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika 61:439–447.

WHITE, G. C. 2000. Population viability analysis: data requirements and essential analyses. Pages 288–331 *in* L. Boitani and T. K. Fuller, editors. Research techniques in animal ecology: controversies and consequences. Columbia University Press, New York, New York, USA.

———, D. R. ANDERSON, K. P. BURNHAM, AND D. L. OTIS. 1982. Capture–recapture and removal methods for sampling closed populations. Los Alamos National Laboratory, Report LA-8787-NERP, Los Alamos, New Mexico, USA.

———, AND K. P. BURNHAM. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study 46 Supplement:120–138.

———, AND R. A. GARROTT. 1990. Analysis of wildlife radio-tracking data. Academic Press, New York, New York, USA.

WILLIAMS, B. K., F. A. JOHNSON, AND K. WILKINS. 1996. Uncertainty and the adaptive management of waterfowl harvests. Journal of Wildlife Management 60:223–232.