

Age-Structured Mark–Recapture Analysis: A Virtual-Population-Analysis-Based Model for Analyzing Age-Structured Capture–Recapture Data

LEWIS G. COGGINS, JR.*

U.S. Geological Survey, Southwest Biological Science Center, Grand Canyon Monitoring and Research Center, 2255 North Gemini Drive, Flagstaff, Arizona 86001, USA

WILLIAM E. PINE III

Department of Fisheries and Aquatic Sciences, University of Florida, 7922 Northwest 71st Street, Gainesville, Florida 32653, USA

CARL J. WALTERS AND STEVEN J. D. MARTELL

Fisheries Centre, University of British Columbia, 2259 Lower Mall, Vancouver, British Columbia V6T 1Z4, Canada

Abstract.—We present a new model to estimate capture probabilities, survival, abundance, and recruitment using traditional Jolly–Seber capture–recapture methods within a standard fisheries virtual population analysis framework. This approach compares the numbers of marked and unmarked fish at age captured in each year of sampling with predictions based on estimated vulnerabilities and abundance in a likelihood function. Recruitment to the earliest age at which fish can be tagged is estimated by using a virtual population analysis method to back-calculate the expected numbers of unmarked fish at risk of capture. By using information from both marked and unmarked animals in a standard fisheries age structure framework, this approach is well suited to the sparse data situations common in long-term capture–recapture programs with variable sampling effort.

Estimating population size is a key component in developing management plans for a wide variety of fisheries. Estimates of population size are often used to evaluate the population status of threatened or endangered species and are a key aspect of most commercial or recreational fisheries stock assessments. The techniques used to estimate population size generally fall into two broad areas, the traditional open- and closed-population capture–recapture models (e.g., Lincoln–Petersen, CAPTURE, Jolly–Seber, etc.; see review by Pine et al. 2003) and age- or size-structured virtual population analysis (VPA)–type methods (Hilborn and Walters 1992). Here we present a new model, called age-structured mark–recapture analysis (ASMR), that combines attributes of both the traditional Jolly–Seber models (Jolly 1965; Seber 1965; Williams et al. 2002) and VPA-type methods widely used in fisheries stock

assessments. We develop this model using data from a long-term tagging program (1989–2002) for humpback chub *Gila cypha* in the Grand Canyon reach of the Colorado River. Explicit details of this tagging program are found in the companion paper to this manuscript (Coggins et al. 2006, this issue).

Overall Model Structure

The ASMR estimation method proposed here is developed in two stages. First, we develop a model for predicting the numbers of marked and unmarked fish at risk of capture over time and age, conditional on survival rate and the marking data. Then we use these predicted numbers at risk to capture along with capture probability parameters to predict the numbers of captures and recaptures that will be observed. After that, we develop a likelihood function for these observations to be maximized by varying the unknown survival, vulnerability, and capture probability or terminal abundance parameters. By using a virtual population analysis method to back-calculate the expected number of unmarked fish at risk of capture, the method avoids treating unmarked fish by age that were alive at the start of sampling and new recruits entering the unmarked population each year as unknown parameters.

The expected numbers of unmarked ($\hat{U}_{a,t}$) and marked ($\hat{M}_{a,t}$) fish by age ($a = 2, \dots, A$) and year ($t = 1, \dots, T$) at risk of capture and recapture are assumed to have varied as follows:

$$\hat{U}_{a+1,t+1} = \hat{S}_a(\hat{U}_{a,t} - m_{a,t}) \quad (1)$$

$$\hat{M}_{a+1,t+1} = \hat{S}_a(\hat{M}_{a,t} + m_{a,t}), \quad (2)$$

where \hat{S}_a is the average annual survival rate of an age- a fish and $m_{a,t}$ is the number of age- a fish marked in year t . Note that equations (1) and (2) assume that

* Corresponding author: lcoggins@usgs.gov

Received May 6, 2005; accepted August 15, 2005
Published online February 3, 2006

every unmarked fish captured received a tag (the same $m_{a,t}$ is used both to decrement \hat{U} and to increment \hat{M}). Note further that multiple recaptures of individual fish within each year are counted as only one recapture event (this model works on annual steps).

To make specific predictions of \hat{U} and \hat{M} for comparison with the data, we must estimate boundary conditions $\hat{M}_{a,1}$ for all a , $\hat{U}_{a,1}$ for all a in year 1, and new recruits $\hat{U}_{2,t}$ for $t > 1$. The initial numbers of marked fish are obviously $\hat{M}_{a,1} = 0$ for all a . There are two choices for dealing with the \hat{U} boundary conditions: (1) “forward propagation,” in which the initial abundance at age and recruitment ($\hat{U}_{a,1}$ and $\hat{U}_{2,t}$) are treated as unknowns to be estimated; and (2) “backward propagation,” in which the oldest age in each year ($\hat{U}_{A,t}$) and all ages in the terminal year ($\hat{U}_{a,T}$) are treated as the unknowns and the other $\hat{U}_{a,t}$ are calculated using the backward recursion or a VPA equation based on solving equation (1) for $\hat{U}_{a,t}$ given $\hat{U}_{a+1,t+1}$, that is,

$$\hat{U}_{a,t} = (\hat{U}_{a+1,t+1} / \hat{S}_a) + m_{a,t} \tag{3}$$

In using the backward propagation approach in ASMR, it is safe to treat $\hat{U}_{A,t}$ as 0 for all t if A is sufficiently beyond the oldest age that fish can attain, and we are left with estimating the number of unmarked fish in the terminal year, $\hat{U}_{a,T}$.

Using the predictions of the numbers at risk of capture from equations (1–3), we then calculate the expected numbers of unmarked and marked fish captured by age and year as follows:

$$\hat{m}_{a,t} = \hat{U}_{a,t} \hat{p}_{a,t} \tag{4}$$

$$\hat{r}_{a,t} = \hat{M}_{a,t} \hat{p}_{a,t}, \tag{5}$$

where $\hat{p}_{a,t}$ is the estimated age- and time-specific capture probability. We assume, conditional on \hat{S}_a and $\hat{p}_{a,t}$, that the observations $m_{a,t}$ and $r_{a,t}$ represent independent samples from Poisson distributions with means given by equations (4) and (5). This is the same as assuming independent binomial sampling of individuals in each \hat{U} - and \hat{M} -age subpopulation with sample capture probability $\hat{p}_{a,t}$. Ignoring terms involving only the data, the Poisson assumption leads to the log-likelihood function

$$\log_e L(m, r | \theta) = \sum_{a=1}^A \sum_{t=1}^T [-\hat{m}_{a,t} + m_{a,t} \log_e(\hat{m}_{a,t})] + \sum_{a=1}^A \sum_{t=2}^T [-\hat{r}_{a,t} + r_{a,t} \log_e(\hat{r}_{a,t})] \tag{6}$$

where θ is the parameter vector to be estimated.

Following Lorenzen (2000), we defined an age-dependent survival function based on a von Bertalanffy

growth function (von Bertalanffy 1938). This formulation allows mortality to decrease with age to a minimum defined by \hat{M}_{adult} , the instantaneous mortality rate suffered by fish that have reached asymptotic length. We obtained an independent estimate of the von Bertalanffy k from growth data presented in the USFWS humpback chub recovery goals (USFWS 2002; see below). The resulting “Lorenzen model” (Lorenzen 2000) allows age-specific survival to be estimated with one unknown parameter, \hat{M}_{adult} , as

$$\hat{S}_a = \left[\frac{e^{k(a+1)} - 1}{e^{ka} - 1} \right]^{\hat{M}_{\text{adult}}/k} \tag{7}$$

The two remaining model specification issues are estimation of the unmarked fish in the terminal year to initialize the back propagation (i.e., $\hat{U}_{a,T}$) and estimation of the age- and time-specific capture probabilities $\hat{p}_{a,t}$. We define three alternative formulations of the ASMR model to incorporate various options.

Specific Models

ASMR 1

In ASMR 1, we calculate the age- and time-specific capture probability as $\hat{p}_{a,t} = \hat{p}_t \hat{v}_{a,t}$, where $\hat{v}_{a,t}$ is the age- and time-specific vulnerability to capture gear and \hat{p}_t is the conditional maximum likelihood estimate of annual capture probability, which is calculated as

$$\hat{p}_t = \frac{\sum_{a=1}^A (m_{a,t} + r_{a,t})}{\sum_{a=1}^A \hat{v}_{a,t} (\hat{U}_{a,t} + \hat{M}_{a,t})} \tag{8}$$

Following standard fisheries virtual population analysis techniques (Hilborn and Walters 1992), we estimate the abundance of unmarked fish in the terminal year to initialize the back propagation as $\hat{U}_{a,T} = m_{a,T} / \hat{p}_{a,T}$. This leaves the parameter vector $\theta = (\hat{v}_{a,t}, \hat{M}_{\text{adult}}, \text{ and } \hat{p}_T)$ to be estimated using nonlinear search routines to maximize equation (6). To further restrict the problem, it is possible to assume that $\hat{v}_{a,t} = 1$ for all ages older than a specified age (i.e., all fish older than a specified age are equally vulnerable to capture). It is also possible to estimate identical $\hat{v}_{a,t}$ schedules for blocks of years that have similar sampling intensities (see Coggins et al. 2006).

ASMR 2

ASMR 2 differs from ASMR 1 only in the initialization of the terminal abundances and in the calculation of the terminal capture probability. Instead of estimating an overall terminal capture probability (\hat{p}_T), we directly estimate terminal abundances ($\hat{U}_{a,T}$). ASMR 2 maximizes equation (6) by varying $\theta = (\hat{v}_{a,t}, \hat{M}_{\text{adult}}, \text{ and } \hat{U}_{a,T})$.

ASMR 3

ASMR 3 differs from formulations 1 and 2 in that more flexibility is allowed in the estimation of age- and time-specific capture probabilities ($\hat{p}_{a,t}$). As in ASMR 2, $\hat{U}_{a,t}$ is estimated directly; however, we use the conditional maximum likelihood estimate for the $\hat{p}_{a,t}$ matrix, that is,

$$\hat{p}_{a,t} = \frac{m_{a,t} + r_{a,t}}{\hat{U}_{a,t} + \hat{M}_{a,t}}. \tag{9}$$

This formulation eliminates the need to specify annual vulnerability schedules at the expense of a much larger parameter set. ASMR 3 maximizes equation (6) by varying $\theta = (\hat{M}_{\text{adult}} \text{ and } \hat{U}_{a,t})$.

One strength of the overall ASMR approach is that it allows the estimation of recruit abundance for years preceding the onset of data collection by using age-specific survival rates and initial-year abundance as follows:

$$\hat{U}_{2,2-a} = \frac{\hat{U}_{a,1}}{\prod_{i=2}^{a-1} \hat{S}_i}. \tag{10}$$

This is an important aspect of the ASMR and a distinction from the “recruitment” parameter estimated by Jolly–Seber type methods (discussed below).

Assignment of Apparent Age at First Capture

We reparameterized the polynomial length-at-age relationship contained in the humpback chub Endangered Species Act recovery goals and based on fish collected in Grand Canyon (USFWS 2002) to a von Bertalanffy growth function necessary for the Lorenzen mortality curve. We then used the inverse von Bertalanffy growth function to assign apparent age as a function of length, that is,

$$a = -\frac{1}{k} \log_e \left(1 - \frac{l}{l_\infty} \right) - a_0, \tag{11}$$

where $k = 0.12$, $a_0 = 0.87$, and $l_\infty = 455$. We conducted sensitivity analyses on the effect of growth parameter misspecification on abundance and recruitment trends.

Model Performance

We tested the estimation methods by generating simulated capture–recapture data from known numbers of fish $\hat{U}_{a,1}$ and $\hat{U}_{2,t}$ from ages 2–30 and years 1989–2002 subject to individual binomial capture, survival, and recapture events over age and time with known $v_{a,t}$ and annual capture probabilities (\hat{p}_t). These tests indicate that the method gives unbiased estimates of the $\hat{p}_t \hat{U}_{a,1}$ up to at least age 10, the $\hat{U}_{2,t}$ for all t , $\hat{v}_{a,t}$ for the time periods mentioned above, and age-specific survival rates \hat{S}_a up to at least $a = 20$. The simulated

recruitment and initial stock estimates $\hat{U}_{2,t}$ and $\hat{U}_{a,1}$ are precise (<10% error) for the period from the early to the mid 1990s and become imprecise for the late 1990s to 2002. Most \hat{S}_a estimates are precise for all a .

We cannot assign an accurate age to each humpback chub at first capture. The recapture size data indicate that growth rates are extremely variable; for example, the average size of an age-10 humpback chub is around 300 mm TL, but fish of this size can be anywhere between about 6 and 15 years of age. When we assign a fish an “age” at first capture, that age may well be predictive of the subsequent size-dependent survival rate (Lorenzen model) but is probably not the correct age for assigning the fish to a cohort. This means that the year-class strengths or apparent cohort sizes $\hat{U}_{a,1}$ and $\hat{U}_{2,t}$ estimated by the procedure outlined above are not the numbers of fish recruiting or initially present by cohort but rather some smoothed or running average of the actual cohort strengths. The estimation method should still be able to detect longer-term trends in abundance and recruitment, but it is less able to detect subtle changes in year-class strength.

The use of VPA back propagation for calculating $\hat{U}_{a,t}$ does not cause the ASMR method to overestimate adult abundance. To evaluate this, we replaced equation (3) with the forward prediction equation (1) for $\hat{U}_{a,t}$ and included the initial unmarked abundances $\hat{U}_{a,1}$ ($a = 2, \dots, 30$) and $\hat{U}_{2,t}$ ($t > 1$) in the set of unknown parameters to be estimated by maximizing equation (6). This resulted in essentially the same abundance and adult mortality estimates as the back propagation method.

While the proposed estimation method is unbiased when supplied with accurate ages at first capture, tests with simulated data that include initial aging error indicate that it produces estimates of \hat{p}_t and \hat{M}_{adult} that are biased downward by about 7% and 11%, respectively. However, it still tracks the simulated recruitment patterns over time and does not create a spurious simulated trend in back-calculated recruitments before the first year of sampling (Figure 1).

As a measure of uncertainty, we sampled the posterior distribution of the parameter estimates using Markov chain–Monte Carlo (MCMC) techniques for each ASMR formulation. Following Gelman et al. (2000), we constructed 95%-credible intervals from MCMC parameter chains of length 1,000 from 200,000 MCMC trials, retaining every 100th trial and disregarding the first half of the chain. We assessed convergence using Gelman and Rubin’s potential scale reduction factor (R Development Core Team 2005).

Discussion

The abundance and mortality trends from the ASMR and Jolly–Seber models are similar (Coggins et al.

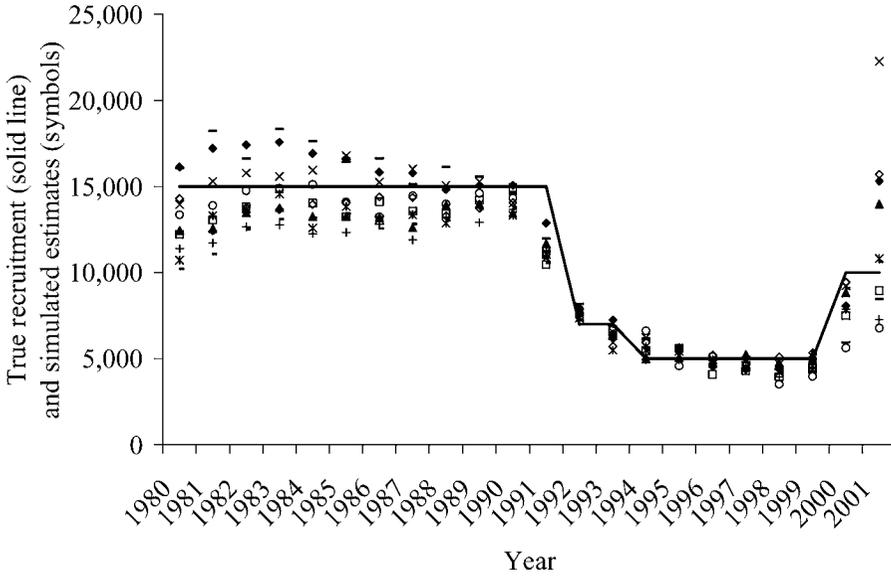


FIGURE 1.—Estimated recruitment over time for simulated data sets assuming stable recruitment before 1989 and error in age assignment due to variation in the simulated growth patterns. Individual growth variation was simulated by setting length at age for each simulated fish i to $l_i(a) = (l_\infty + d_i)(1 - e^{-k(i + 0.42)})$, with the deviations in asymptotic length d_i normally distributed with mean zero and standard deviation 30. Each symbol type represents results for a different simulated data set incorporating stochastic error in age assignment.

2006), although the two approaches use distinct methods to derive the parameter estimates. The ASMR model predicts the number of both unmarked and marked fish available for capture using marking data and survival estimates. This reconstructed annual abundance at age is then used along with age- and time-specific capture probabilities to predict the number of both marked and unmarked fish captured during each sampling effort. Finally, predicted captures are compared with the observed capture data to estimate model parameters (i.e., survival and capture probabilities or terminal abundance). This approach differs from Jolly–Seber models, which primarily rely on recaptures of previously tagged individuals for survival and population size estimation. While Jolly–Seber models can be parameterized to include both age- and time-dependent factors (i.e., “Jolly–age” models; Pollock 1981), we have found that capture–recapture data sets of long-lived species with many age-classes (30 in this example with humpback chubs) and low capture probabilities across ages and years (generally <0.2) often contain many years with few or no individuals in several year-classes. In general, we have found that age-structured Jolly–Seber models do not perform well in these situations without parameter constraints because of obviously sparse data. Because of the large amount of additional age-structure in-

formation and assumptions built into the ASMR model related to both the tagged and untagged animals, ASMR models may fit sparse data situations better than unconstrained Jolly–Seber models if ASMR model assumptions are met. However, as with traditional VPA models, ASMR abundance estimates can become unstable with low overall capture probabilities.

A primary purpose of ASMR is to evaluate the recruitment responses of individual year-classes in response to adaptive management experiments related to water manipulation and exotic species removal. ASMR differs from Jolly–Seber methods in estimating recruitment by reconstructing year-classes based on age-specific survival rates and initial abundances. In Jolly–Seber methods, “recruitment” into an age-class is a combination of immigrants and survivors from the previous time period; thus recruitment can occur with each year-class. With ASMR, recruitment is only allowed into the first year-class by design.

ASMR assumes that the relationship between age and survival is governed by the size of the animals, as described by Lorenzen (2000). We examined this assumption within ASMR by individually estimating each age-specific survival rate and found good agreement with the Lorenzen function in the overall shape and magnitude of the survival rates. However, incorporating the Lorenzen function provides a sub-

stantial benefit by shrinking the size of the parameter set. Though it is theoretically possible to introduce this functional form into a traditional Jolly–Seber model, we find the ASMR method more flexible in accommodating these types of fisheries-specific model structure for evaluating hypotheses related to various functional forms. Additionally, the incorporation of a biologically reasonable assumption related to the effect of size on natural survival rate probably allows ASMR models to interpret periods of sparse data more efficiently than traditional Jolly–Seber models.

Two other important methodological differences exist between ASMR and Jolly–Seber-type methods. First, ASMR uses “summary”-type statistics of captures and recaptures as opposed to the individual-capture-history approach used in applications such as MARK. We acknowledge that the individual-capture-history approach may provide some additional information on survival and capture probability (Nichols and Pollock 1983) and facilitate the use of individual covariates such as length. Future formulations of ASMR models will examine the use of individual capture histories.

A second key difference between the two methods is the use of a Poisson distribution to estimate the number of captures and recaptures in the ASMR method in contrast to the multinomial approach used in Jolly–Seber methods. Binomial distributions can be modeled as a series of independent Poisson distributions, both leading to the same maximum likelihood estimates (Sandland and Cormack 1984). The use of a Poisson distribution may lead to estimates of population size that have a slightly lower variance, but the difference is probably very small (C. Schwarz, Simon Fraser University, personal communication). The use of independent Poisson distributions to model recaptures is slightly different, as the same fish could be recaptured multiple times (i.e., the Poisson distributions are not independent). However, the likelihoods used for the recaptures do approximate generalized estimating equations, where Poisson distributions are commonly used when modeling counts. A drawback to the ASMR approach is that this routine does not easily lend itself to routine statistical model selection procedures (e.g., the likelihood ratio test or Akaike information criterion, as used in MARK). This is because the fitting routines employed are a combination of Poisson likelihood functions and relatively simple estimating equations. The model selection criteria assume that the estimating functions are pure likelihoods and not a combination approach as used here (C. Schwarz, Simon Fraser University, personal communication). Future work with ASMR models

should include exploring appropriate model selection procedures.

Acknowledgments

We acknowledge the comments of C. Schwarz regarding the differences between Poisson and multinomial distributions in capture–recapture analysis.

References

- Coggins, L. G., Jr., W. E. Pine III, C. J. Walters, D. R. Van Haverbeke, D. Ward, and H. C. Johnstone. 2006. Abundance trends and status of the Little Colorado River population of humpback chub. *North American Journal of Fisheries Management* 26:76–88.
- Gelman, A. B., J. S. Carlin, H. S. Stern, and D. B. Rubin. 2000. *Bayesian data analysis*. Chapman and Hall, New York.
- Hilborn, R., and C. J. Walters. 1992. *Quantitative fisheries stock assessment: choice, dynamics, and uncertainty*. Chapman and Hall, New York.
- Jolly, G. M. 1965. Explicit estimates from capture–recapture data with both death and immigration–stochastic model. *Biometrika* 52:225–247.
- Lorenzen, K. 2000. Allometry of natural mortality as a basis for assessing optimal release size in fish-stocking programmes. *Canadian Journal of Fisheries and Aquatic Sciences* 57:2374–2381.
- Nichols, J. D., and K. H. Pollock. 1983. Estimation methodology in contemporary small mammal capture–recapture studies. *Journal of Mammalogy* 64:253–260.
- Pine, W. E. III, K. H. Pollock, J. E. Hightower, T. J. Kwak, and J. A. Rice. 2003. A review of tagging methods for estimating fish population size and components of mortality. *Fisheries* 28(10):10–23.
- Pollock, K. H. 1981. Capture–recapture models allowing for age-dependent survival and capture rates. *Biometrics* 37:521–529.
- R Development Core Team. 2005. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available: <http://www.R-project.org>. (June 2005)
- Sandland, R. L., and R. M. Cormack. 1984. Statistical inferences for Poisson and multinomial models for capture–recapture experiments. *Biometrika* 71:27–33.
- Seber, G. A. F. 1965. A note on the multiple-recapture census. *Biometrika* 52:249–259.
- USFWS (U.S. Fish and Wildlife Service). 2002. *Humpback chub (Gila cypha) recovery goals: amendment and supplementation to the Humpback Chub Recovery Plan*. USFWS, Denver, Colorado. Available: www.gcmrc.gov/library/reports/biological/Fish_studies/USFWS/Humpbackchub_recoverygoals2002.pdf. (May 2005).
- von Bertalanffy, L. 1938. A quantitative theory of organic growth (inquiries on growth laws II). *Human Biology* 10:181–213.
- Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. *Analysis and management of animal populations: modeling, estimation, and decision making*. Academic Press, San Diego, California.