

## EVALUATION OF CLOSED CAPTURE–RECAPTURE METHODS TO ESTIMATE ABUNDANCE OF HAWAIIAN MONK SEALS

JASON D. BAKER<sup>1</sup>

Honolulu Laboratory, Pacific Islands Fisheries Science Center, National Marine Fisheries Service, NOAA,  
2570 Dole Street, Honolulu, Hawaii 96822-2396 USA, and  
University of Aberdeen, School of Biological Sciences, Lighthouse Field Station, Cromarty, Ross-shire IV11 8YJ UK

**Abstract.** Numerous capture–recapture methods have been developed to estimate abundance, yet the performance of these models is only rarely judged by comparison with true abundance. This study presents a rare opportunity to assess capture–recapture estimates in a free-ranging population with known minimum abundance. Hawaiian monk seal abundance historically has been characterized using a trend index or has been estimated using simple enumeration. Here, I evaluate the applicability of various closed-population capture–recapture models to estimating Hawaiian monk seal abundance and its associated error. I analyzed 12 data sets (two years from each of six subpopulations) representing a wide variety of sampling and logistical scenarios, using models that explored the effects of animal size class (juvenile, subadult, or adult), tag status, and sighting location on initial capture and recapture probabilities. I also explored various models to account for capture heterogeneity among individuals. Size and sex effects always significantly improved model fits, and tag status and location effects were also frequently influential. In most cases, abundance estimated from capture–recapture models was substantially lower than known minimum abundance, suggesting the influence of individual capture heterogeneity. Attributes of individuals known to be alive, but not captured during systematic surveys, did not reveal patterns that explained sources of capture heterogeneity. In some cases, mixture models produced estimates that were less biased but were associated with very large confidence intervals. Among the model types examined, those available in Program CAPTURE performed best; although they are still prone to negative bias, these models nevertheless may prove useful in characterizing population trends in Hawaiian monk seals. This study demonstrates that selection of appropriate closed capture–recapture models can be substantially improved by independent validation.

**Key words:** abundance estimation; bias; capture heterogeneity; capture–recapture; Hawaiian monk seal; *Monachus schauinslandi*; validation studies.

### INTRODUCTION

Precise and unbiased estimates of abundance are a fundamental requirement when managing exploited or endangered populations. Both applied and natural marks have long been used to facilitate individually based behavioral and demographic studies. This has led to the development of sophisticated analytical techniques that use sightings of identified individuals within a capture–recapture framework to estimate abundance (e.g., Otis et al. 1978, Seber 1982, Schwarz and Seber 1999). These studies require consideration of the potential for, and ramifications of, violating various assumptions of capture–recapture methods, including population closure, loss of marks, and heterogeneity in individual behavior leading to unequal capture probabilities (e.g., Otis et al. 1978). Although more recent

models have been designed to account for such capture heterogeneity, their success in reducing negative bias in abundance estimates can be difficult to determine. Only rarely have capture–recapture estimates been tested against known populations sizes (Edwards and Eberhardt 1967, Strandgaard 1967, Mares et al., 1981, Greenwood et al. 1985, Boulanger and Krebs 1994, Manning et al. 1995, Koper and Brooks 1998, Cross and Waser 2000), and their performance when applied to wider ranging or socially complex mammals remains unclear. The consequences of bias in abundance estimates for management of such populations could be severe. Inadequate consideration of positive bias could lead to over-exploitation in harvested species. Alternatively, negative bias might result in economic consequences for industries such as fisheries, where take limits can be influenced by the abundance of endangered populations.

Potential to explore these biases is likely to be greatest in endangered species, where low numbers and restricted range provide opportunities for alternative techniques to provide comparative data on abundance. The Hawaiian monk seal (*Monachus schauinslandi*) is

Manuscript received 25 April 2003; revised 13 October 2003; accepted 9 November 2003; final version received 8 December 2003. Corresponding Editor: P. K. Dayton.

<sup>1</sup> Address for correspondence: Honolulu Laboratory, Pacific Islands Fisheries Science Center, National Marine Fisheries Service, NOAA, 2570 Dole Street, Honolulu, Hawaii 96822-2396 USA. E-mail: jason.baker@noaa.gov

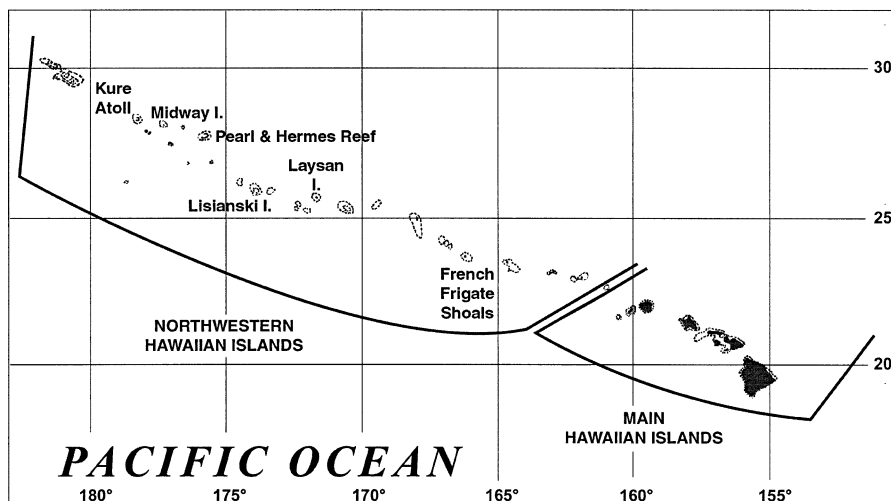


FIG. 1. Location of the six main subpopulations of Hawaiian monk seals in the northwestern Hawaiian Islands, USA.

one such endangered species that primarily inhabits the remote northwestern Hawaiian Islands (NWHI), where ~1400 seals reside predominately in six main subpopulations (Ragen and Lavigne 1999, Carretta et al. 2002). Early efforts to characterize abundance involved using counts on land as an index of population size (Kenyon and Rice 1959), and steep declines in these counts led to the species being listed as endangered under the U.S. Endangered Species Act in 1976 (U.S. Department of Commerce 1976a, b).

Subsequent efforts to monitor and foster the recovery of Hawaiian monk seals have resulted in an extensive research program in the six main NWHI subpopulations for more than two decades. In recent years, the monitoring effort has been so intensive that almost all individual seals using most of these sites are likely to be identified through resights of tags or by recognition of permanent natural marks or applied temporary pelage bleach marks. This enumeration process provides a known minimum abundance (Carretta et al. 2002), but it is a point estimate lacking an estimate of associated error, such that the potential number of uncounted seals cannot be assessed. Because sightings of marked seals are collected in a systematic manner, on repeated censuses through each season, the application of capture–recapture techniques to these data could overcome this problem and provide both a point estimate and an estimate of error.

In this paper, I investigate whether applying capture–recapture models to existing Hawaiian monk seal data can generate abundance estimates that represent an improvement over both the beach count index and simple enumeration. This database of individually based sightings presents a rare opportunity to independently evaluate the capture–recapture abundance estimates, specifically with respect to the influence of capture heterogeneity among individuals. The six monk seal populations studied vary in terms of their physiography,

population size, and logistic challenges that constrain sampling effort. Thus, they represent a diversity of scenarios in which the performance of the capture–recapture models can be evaluated. Furthermore, seals were identified both during systematic censuses (which conformed well to capture–recapture methods) and during nonsystematic and opportunistic sightings (which did not). Thus, some seals were known to be present in the population, even though they were not seen on any systematic census. The total number of individuals identified during both types of sampling constitutes a known minimum population size, which is compared to estimates derived from capture–recapture analysis using only the systematic data.

#### METHODS

Capture–recapture data analyzed here were extracted from the Hawaiian monk seal sighting databases from six locations: French Frigate Shoals, Laysan Island, Lisianski Island, Pearl and Hermes Reef, Midway Atoll, and Kure Atoll (Fig. 1). The fundamental methods employed were consistent among all sites and years. However, varying physiography among the sites and logistical constraints led to variability in the timing, duration, and frequency of sampling occasions. General methods will be described first, followed by salient site-specific characteristics. Because unequal capture probabilities tend to result in biased abundance estimates (Otis et al. 1978), efforts were made to sample available data in such a way as to minimize known sources of heterogeneity introduced through the design of resighting studies.

#### Marking

Since the early 1980s, all or nearly all pups have been marked on each hind flipper with a unique colored plastic tag. Many older animals, which either had never been tagged or whose tags required replacement, also

have been captured and tagged opportunistically. As a result, the majority of seals bear tags. In addition, photographic and hand-drawn records of natural markings (shark bite and other scars, permanent natural pelage patterns) have been collected and updated annually for all seals. Finally, to facilitate resighting seals and to provide a unique mark for seals without tags or distinguishing natural marks, most seals have been given temporary pelage bleach marks.

#### *Recaptures*

The presence of individual seals was recorded both during complete systematic surveys (hereafter referred to as censuses) of all landing areas and during incidental sightings or partial censuses of beaches (hereafter referred to as noncensus sightings). For capture-recapture analysis, individual capture histories were constructed only from complete censuses. Censuses were conducted by walking the entire perimeter of all islands within an atoll and systematically searching for and identifying as many seals encountered as possible. This tended to equalize the probability that individuals would be identified, given that they were present on land. Seals seen in the water were excluded from the capture histories, because the identity of these seals could be less certain and capture probability would vary depending on the nature of the distinguishing mark (e.g., bleach mark vs. flipper tag).

During censuses, seal identity, sex, size class (pup, juvenile, subadult, and adult size; see Stone [1984]), and location were recorded. Censuses at multiple-islet atolls were completed in two days or less. Censuses of single-island subpopulations began at 1300 hours and were always completed within the day.

Some animals counted on censuses were not identified. Care was taken to avoid disturbing seals, such that alert and aware animals were not always approached closely enough to identify. Also, even approachable seals sometimes could not be identified on each sighting. The number of seals present but not identifiable on a given census was recorded.

Noncensus sightings were collected on surveys that did not cover all island perimeters, that were focused on particular age and sex groups, or that were simply sightings recorded incidental to other activities. These sightings were excluded from capture-recapture analysis because their inclusion would certainly introduce capture heterogeneity through unequal effort. For example, on multiple-islet atolls, seals that frequented the islet where researchers were based were more likely to be resighted than those that favored distant islets, which researchers visited less often. Likewise, nursing mothers were more closely monitored than other sex and age groups, giving them higher noncensus capture probabilities.

The duration of field seasons and the number of censuses completed per season varied among sites and over time, although they usually occurred during the main

period of pupping and breeding (March–August, Johanos et al. [1994]). Atoll physiography, weather conditions, and other factors variably influenced the census schedule. Conducting censuses was but one of many research objectives with which field teams were charged. Often, researchers were at their field sites for several weeks or more prior to conducting the first census and also remained after the final census. Thus, the census period constituted a subset of the full field season, and noncensus sightings occurred both prior to, and after, the census period. Site-specific details of field methods will be discussed.

#### *Single islands: Laysan and Lisianski*

Laysan and Lisianski are relatively large (11 and 5 km perimeter, respectively) islands, where field study seasons were ~5 months long. All locations where seals land were readily accessible to researchers each day, and weather conditions rarely affected the schedule of surveys. Typically, Laysan and Lisianski field teams spent approximately the first month of each field season identifying seals and applying bleach marks to most of the population. Subsequently, they began conducting censuses every few days, usually completing 20–30 censuses per season. During the past decade, the number of seals (excluding pups) identified at these sites has ranged from 206 to 272 at Laysan and 143 to 197 at Lisianski. Long before the end of each field season, sightings of seals not yet seen that year almost completely ceased, suggesting that the entire population had been identified.

#### *Multiple-islet Atolls*

The remaining four sites consist of coral atolls of varying size, with varying numbers of islets. Seals frequently move among the islets within atolls. Researchers were stationed at one islet at each atoll, and traveled by small boat to the various islets to conduct census- and noncensus-related research. Islets range from substantial vegetated islands to small sandspits only a few tens of meters in diameter. Ability to reach all the islets was highly dependent on weather, sea state, and distances between islets.

*French Frigate Shoals.*—This largest of the NWHI atolls consists of a 36 km diameter lagoon enclosed by a fringing reef and dotted with nine islets where seals land. Weather and sea conditions had to be favorable to achieve a full census of all islets within two days. Because the population is relatively large (240–351 nonpups identified per year during the past decade) and spread among the variably accessible islets, little bleach marking has been done, making repeat identification within a season more difficult. This site recently has harbored the largest subpopulation of Hawaiian monk seals, which has steeply declined for more than a decade (Craig and Ragen 1999, Carretta et al. 2002). Field seasons have tended to be 4–5 months

long, typically with 8–10 censuses conducted per season.

*Pearl and Hermes Reef.*—This large atoll (diameter 32 km) has eight islets and a monk seal population that has been growing since the early 1970s (the number of nonpups identified has ranged from 198 to 227 per year during the past decade). As on French Frigate Shoals, conducting censuses at Pearl and Hermes Reef is highly weather dependent. Research at this atoll typically occurs over a 7–10 week period, within which 8–10 censuses are conducted.

*Midway atoll.*—This relatively small (11 km in diameter) atoll, quite tractable for monitoring, consists of two primary islands and one small spit, all within 2 km of one another. The monk seal population here has grown from just a few individuals in the 1980s to ~50–60 animals (excluding pups) recently. Very little research effort was expended until the population began to grow, after which the field seasons extended from 9 to 12 months. Up to 50 censuses have been conducted in some years.

*Kure atoll.*—This small atoll (9 km diameter) has only two permanent islands. In recent years, field season durations usually have been 9–12 weeks, during which 10–15 censuses are completed. The number of seals older than pups identified at this site has increased from 64 in 1990 (Van Toorenburg et al. 1993) to ~100 in recent years.

#### Data analysis

Data sets from two years from each of the six subpopulations were selected for analysis. Years representing “best case” scenarios were chosen with the rationale that if the results were not satisfactory, less ideal data sets would yield no better results. Several elements constituted best case data sets. First, relatively large numbers of censuses were desirable to reduce the error of abundance estimates. Further, because monk seals alternate time on land with trips to sea, it was desirable to have censuses relatively evenly spread over most of the field season so that each animal in the population would probably be available for capture on several occasions. To further minimize capture heterogeneity, years with a relatively low proportion of unidentifiable seals encountered per census (i.e., when the population was well marked), and low variability in total number of seals identified per census, were chosen.

Capture–recapture histories from censuses were analyzed using closed-population maximum likelihood estimators (MLE) in Program MARK (White 2002) and various other estimators available through Program CAPTURE (Otis et al. 1978, White et al. 1982, Rexstad and Burnham 1991). Open models using multiple capture events both within and between consecutive years (i.e., Kendall et al. 1995) were not appropriate because some individuals had temporary bleach marks not maintained between years. The first time that a seal

was identified during a field season was treated as the initial capture (analogous to marking), and any subsequent resightings were treated as recaptures. Variables thought to potentially affect capture probabilities were used to subdivide populations into strata (“groups” in Program MARK parlance), the parameters of which were fitted separately. Models were used to explore effects of size class (adult, subadult, or juvenile), sex, and tag status (presence or absence of flipper tags). At multi-islet atolls, geography might have influenced capture probability; thus, the islet of initial sighting was used as a proxy location effect to further distinguish groups. Time dependence (capture occasion) of capture probabilities was also explored.

When the number of individuals in a group (i.e., a particular sex, size, and location combination) was either zero or very low, nonsensical parameter estimates with enormous variances resulted. In such cases, individual islet locations were collapsed until adequate cell counts were achieved. To the degree possible, adjacent islets with similar physiography were combined. Most untagged animals were either adults that had never been tagged or those that had lost their tags. The effect of tag status therefore was analyzed only for adults. When all variables of interest (size, sex, location, and adult tag status) were considered simultaneously, the number of animals in one or more groups was typically too low to obtain identifiable parameter estimates. To avoid this problem, separate analyses were done for each data set; one involving location, sex, and size and the other exploring adult tag status, sex, and size.

Fundamental assumptions for capture–recapture abundance estimation are presented in Otis et al. (1978). Populations were assumed to be closed during the study, i.e., there were no additions (births or immigration) or losses (mortality or emigration) during a field season. Because censuses were conducted within the periods when most births occur, only nonpup abundance was estimated. Field seasons were of sufficient duration that mortality could have occurred during the census period. Likewise, some movement between subpopulations occurs, which constitutes emigration or immigration (Harting 2002). Finally, seals alternate time ashore with time at sea, a form of temporary emigration. The implications of violating the closed-population assumption will be discussed.

Other assumptions were that marked animals did not lose their marks during the study and that all marks were correctly recorded (Otis et al. 1978). These assumptions were well met. The redundant identification system involving tags, bleach marks, and natural marks makes it very unlikely that an animal would “lose its mark” during a season. Further, data were collected on standardized forms, which allow for notation of any uncertainty in the field identification. Data were checked manually, and quality control checks involving cross-referencing of tag numbers, bleach marks,

size, and sex of animals were employed to find mis-identifications. Only sightings recorded as absolutely certain were used in this study.

Simple closed-capture models in Program MARK assume no capture heterogeneity within groups. Capture heterogeneity was explored using mixture models, in which initial capture probability and recapture probability parameters are modeled as mixtures of values (Norris and Pollock 1995, Pledger 1998, 2000). This allowed for estimated proportions of groups to be subject to different capture and recapture probabilities independent of any identified characteristic (i.e., sex, size, location). Additional estimators for treating capture heterogeneity were analyzed using Program CAPTURE.

Models in Program MARK were fit beginning with a saturated closed-capture model (no individual heterogeneity) including all variables of interest and their interactions. Subsequently, terms that appeared to be insignificant based upon examination of parameter estimates were eliminated one at a time, beginning with interactions and followed by main effects. The results were ranked and evaluated using the small-sample Akaike's information criterion ( $AIC_c$ ; see Anderson et al. [2000]). The total abundance estimate from the model with the lowest  $AIC_c$  was considered best. In cases in which two or more models were roughly equally supported according to  $AIC_c$ , parsimonious models (those with fewer estimated parameters) were preferred (Burnham and Anderson 1998).

Next, the best-fit model for each data set was reparameterized to allow for two mixtures within each group in order to further explore unequal capture probabilities. A "null" mixture model was also fit with two mixtures and no group distinctions.

Additional abundance estimates were generated using the model selection procedure available in Program CAPTURE (Otis et al. 1978, Rexstad and Burnham 1991), whereby a testing algorithm determines which of a suite of models involving combinations of time, behavior, and heterogeneity effects is most appropriate for a given set of capture histories. Estimates were obtained using the model type assigned the highest rank by the algorithm. When the chosen model included all three effect types ( $M_{tbh}$ ), which has no estimator in Program CAPTURE, the second highest rank model was used. When two estimators of the indicated model type were available, both were run.

Parameters estimated in closed-capture models are discussed using the following nomenclature:  $p_i$  is estimated initial capture probability for group  $i$ ;  $c_i$  is estimated recapture probability for group  $i$ ;  $N_i$  is estimated abundance of group  $i$ ;  $M_i$  is the total number of unique individuals in group  $i$  captured during all capture events combined; and  $f_{0,i}$  is the estimated number of individuals of group  $i$  not seen during any capture event, such that:

$$N_i = M_i + f_{0,i}.$$

For this study,  $N$  was of primary interest. Probabilities  $p$  and  $c$  drove the estimation of  $N$ , but were of little interest. Separate estimates of  $N$  are obtained for each of the  $i$  modeled groups. The  $N_i$  are the sum of  $M_i$  and  $f_{0,i}$ , with the  $f_{0,i}$  fitted using the log-link function. The  $f_{0,i}$  on the log scale are referred to as  $\beta_i$ , and Program MARK provides variance estimates for the  $\beta_i$ . Estimated variances for the real-scale  $f_{0,i} = e^{\beta_i}$  were calculated using the delta method (Seber 1982). Thus,

$$\text{var}(e^{\beta_i}) = \left( \frac{\partial e^{\beta_i}}{\partial \beta} \right)^2 \text{var}(\beta_i)$$

or

$$\text{var}(e^{\beta_i}) = (e^{\beta_i})^2 \text{var}(\beta_i).$$

Expanding this calculation to the variance of the total estimated abundance,  $\sum N_i$ , the row vector of  $e^{\beta_i}$  was multiplied by the variance-covariance matrix of  $\beta_i$  values, and the product was post-multiplied by the column vector of the  $e^{\beta_i}$  values:

$$\text{var}\left(\sum N_i\right) = [e^{\beta_i}]^T \cdot [\text{var cov}] \cdot [e^{\beta_i}].$$

Having obtained estimates of total abundance and its variance, I calculated 95% confidence intervals following Rexstad and Burnham (1991).

Abundance estimates from systematic censuses were compared with the total number of seals identified from all (including noncensus) sightings within a season. Discrepancies between the two could occur if some animals were present only during the portion of the season when censuses were not conducted. The census period was defined as the dates between the first and last census in a given field season. Sightings of seals never seen during any census were examined to determine whether they suggested that the populations were not closed during the entire field season. For each data set, these seals were categorized as seen only before the first census, seen only after the last census, seen only before and after the census period, and seen at least once during the census period (although only during a noncensus sighting).

To investigate possible sources of capture heterogeneity, I used logistic regression analysis of the binary response "seen" or "not seen" during censuses to investigate whether size, sex, location, adult tag status, and interactions among these variables significantly influenced which seals were never seen during census. Each of the 12 data sets was analyzed separately because differing site physiography and year-specific patterns in sighting effort could influence individual capture heterogeneity. Data from the same site in different years were not combined, because many individuals were present in both years, and their observations would likely be non-independent. Fisher's exact test was used when the number of seals not seen on census was sufficiently small that expected values in simple

TABLE 1. Data sets selected for capture–recapture estimation of monk seal subpopulation abundance.

Subpopulation	Year	Site description	Field effort (d)	Number of censuses	Size	Sex	Tag status	Location
Lisianski	1997	single island	102	18	X	X		NA
Lisianski	1999	single island	128	28	X	X		NA
Laysan	2000	single island	149	24	X	X	X	NA
Laysan	1995	single island	89	13	X	X		NA
Midway†	1999	three-islet atoll	365	50	NA	NA	NA	NA
Midway‡	2000	three-islet atoll	258	27	X	X	NA	NA
Kure	1991	two-islet atoll	170	44	X	X	X	X
Kure	2000	two-islet atoll	73	13	X	X		X
French Frigate Shoals	1998	nine-islet atoll	127	10	X	X	X	
French Frigate Shoals	1999	nine-islet atoll	253	10	X	X	X	X
Pearl and Hermes Reef§	1999	eight-islet atoll	50	8	X	X	X	NA
Pearl and Hermes Reef	2000	eight-islet atoll	69	10	X	X		X

Note: An “X” indicates that the factor in the column heading was found to significantly influence either initial capture probability (*p*), recapture probability (*c*), or both in Program MARK closed-capture models; NA means not applicable.

† All groups were combined (any subdivision of groups precluded model convergence).

‡ Only sex and size groups were analyzed (distinction of tag status and location groups precluded model convergence).

§ Only sex, size, and tag status were analyzed (location could not be analyzed due to low sample sizes in certain size–sex–location groups).

|| At Pearl and Hermes Reef, there are seven permanent islets and a small number of scattered ephemeral sandspits sometimes used by seals.

contingency tables were less than 1, or when 20% of expected values were less than 5 (Zar 1984).

RESULTS

Data sets

Various scenarios are represented in the selected data sets (Table 1, Fig. 2). These include single-island subpopulations (Laysan and Lisianski) with 3–5 month long field seasons and a large number of censuses (capture occasions); small atolls (Midway and Kure), usually with long seasons and many censuses; and large, complex atolls with numerous islets (French Frigate Shoals and Pearl and Hermes Reef) and either short or

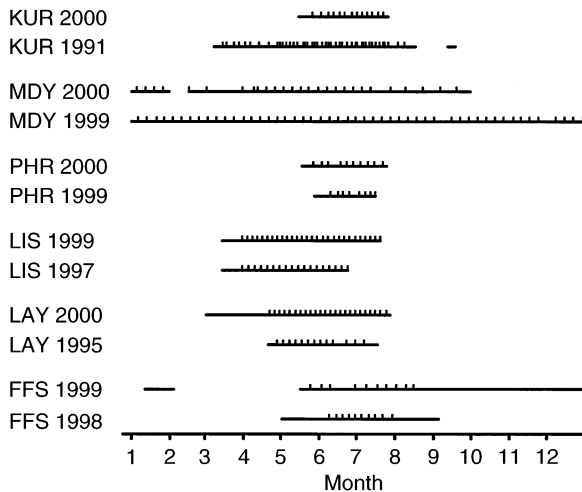


FIG. 2. Duration of field effort (solid lines) and timing of systematic census (vertical marks) of Hawaiian monk seals for 12 analyzed data sets. Site abbreviations are: PHR, Pearl and Hermes Reef; FFS, French Frigate Shoals; KUR, Kure Atoll; MDY, Midway Atoll; LAY, Laysan Island; and LIS, Lisianski Island. Month 1 = January.

long field seasons, but a relatively low number of censuses. The number and distribution of censuses over time within the total period of field effort also varied (Fig. 2).

Best-fit models from Program MARK

The parameters found to significantly influence *p* and *c* varied substantially among the data sets (Table 1); however, some patterns emerged. First, models with time-dependent capture probabilities had the worst AIC<sub>c</sub> values among all models fit for all data sets. A second trend was that for all but one of the 12 data sets, models with distinct parameters for initial capture (*p*) and recapture (*c*) probabilities performed better than those with equal initial and recapture probabilities. However, there was no discernible pattern in these differences. In some data sets, *c*'s tended to be greater than *p*'s; in others, the opposite was found. Even within data sets there was variability such that some size or sex groups had higher *p*'s and others had higher *c*'s at the same location and year.

Accounting for size and sex improved fits to every data set (Table 1). This did not necessarily mean that each size (juvenile, subadult, adult) and sex combination had different capture probabilities, rather that at least one size or sex group differed from the others. In half of the data sets, tagged animals had significantly higher estimated capture probabilities. Finally, for most of the multi-islet atoll data sets, first capture location influenced capture or recapture probabilities for some groups of animals, although not consistently. That is, there were often size × location and sex × location interactions in capture probabilities.

For each data set, numerous models with a wide variety of parameters were evaluated, and the quality of the fits differed greatly as assessed by AIC<sub>c</sub>. How-

TABLE 2. Sample of several models fitted to the French Frigate Shoals 1998 monk seal capture histories using Program MARK.

Model parameters		AIC <sub>c</sub>	ΔAIC <sub>c</sub>	N
Initial capture probability, <i>p</i>	Recapture probability, <i>c</i>			
Adult males vs all others	adult males vs. all others	1702.06	0	279.8
Size × sex	size × sex	1712.00	9.94	280.0
Size × sex	size × sex × location two-way interactions only	1714.73	12.67	280.0
Size × sex × location two-way interactions only	size × sex × location two-way interactions only	1722.43	20.37	280.8
Size × sex × location with three-way interactions	size × sex × location with three-way interactions	1726.23	24.17	281.6
Time dependence and additive size, sex, location effects	time dependence and additive size, sex, location effects	1861.13	159.07	267.0

*Notes:* Groups consisting of combinations of size (juvenile, subadult, or adult), sex, and two within-atoll location categories were examined. The multiplier “×” denotes all additive and interaction terms. AIC<sub>c</sub> and ΔAIC<sub>c</sub> are ranked with the best model on top. Although model performance based upon AIC<sub>c</sub> varied widely, with a simple model being clearly best, the estimated total abundance (*N*) varied little.

ever, each model tended to produce very nearly exactly the same estimate of total abundance. Thus, one could definitively determine which of a suite of models fit best, but this selective process had no real bearing on the parameter of ultimate interest, *N*. (A typical example is presented in Table 2.)

#### *Abundance estimates*

Total estimated abundances from all types of capture–recapture models examined were compared to the total number of seals identified during systematic census and the total number identified during both census and noncensus observations (Fig. 3). In the majority of cases, capture–recapture analysis underestimated true abundance, and the estimated population size was equal to, or just slightly larger than, the total number seen on census. This was typically much lower than the total number of individual seals known to have been present when noncensus sightings were included. In about half of the Program MARK models, the upper 95% confidence limits for total abundance were lower than the known minimum abundance (Fig. 3).

In particular, analysis of the single-island and small-atoll subpopulations (Laysan, Lisianski, Midway, and Kure) produced negatively biased estimates of abundance. Also, because of the large number of censuses conducted at these sites, the estimated variance of abundance was small, often essentially zero. Mixture models did not correct the negative bias in abundance for these data sets; instead, they tended to produce larger variance estimates.

Models fit to the data from large multi-islet atolls with relatively few (8–10) censuses produced somewhat varying results (Fig. 3). At French Frigate Shoals, the simple closed-capture models produced negatively biased abundance estimates. Several of the mixture models produced point estimates greater than the known minimum abundance, and all had upper confidence limits exceeding that number. However, the width of the confidence intervals was often very large

(e.g., French Frigate Shoals, 1998, in Fig. 3). Estimates for Pearl and Hermes Reef differed in that most models produced point estimates of abundance that exceeded known minimum abundance. Again, many of the estimates were imprecise, with extremely high upper confidence limits.

Program CAPTURE tended to produce more useful results. Estimates were frequently less negatively biased and confidence intervals were, by and large, better behaved (i.e., neither unrealistically small nor too large to be practically useful). Notably, the algorithm always selected model types that included heterogeneity effects. These models and their usual abbreviations in the capture–recapture literature include two involving individual heterogeneity (jackknife estimator  $M_{th}$ ; Burnham and Overton [1978, 1979]; and Chao’s  $M_{th}$ , Chao [1989]), one with time and heterogeneity effects ( $M_{th}$ ; Chao et al. [1992]), and two with behavioral response to capture and heterogeneity (removal  $M_{bh}$ ; Otis et al. [1978]; and Pollock and Otto [1983]  $M_{bh}$ ).

#### *“Uncaptured” individuals*

With the exception of those at French Frigate Shoals, most seals sighted only during noncensus activities were seen at least once between the first and final census (Fig. 4). Thus, they were alive and present at least some of the time when systematic censuses were being conducted, but they were not captured during those events. In some cases the census period encompassed virtually the entire field season (e.g., at Midway Atoll), leaving little opportunity for sighting outside that period. In other cases there was significant nonsystematic sighting effort before or after the census period (e.g., French Frigate Shoals, Fig. 2). Even at French Frigate Shoals, most of the seals missed on census were seen during, after, or only before and after the census window. These seals were alive during the census period, but some may not have been present at the site during the systematic surveys. The few seals seen only before the census period might have died prior to the census period.

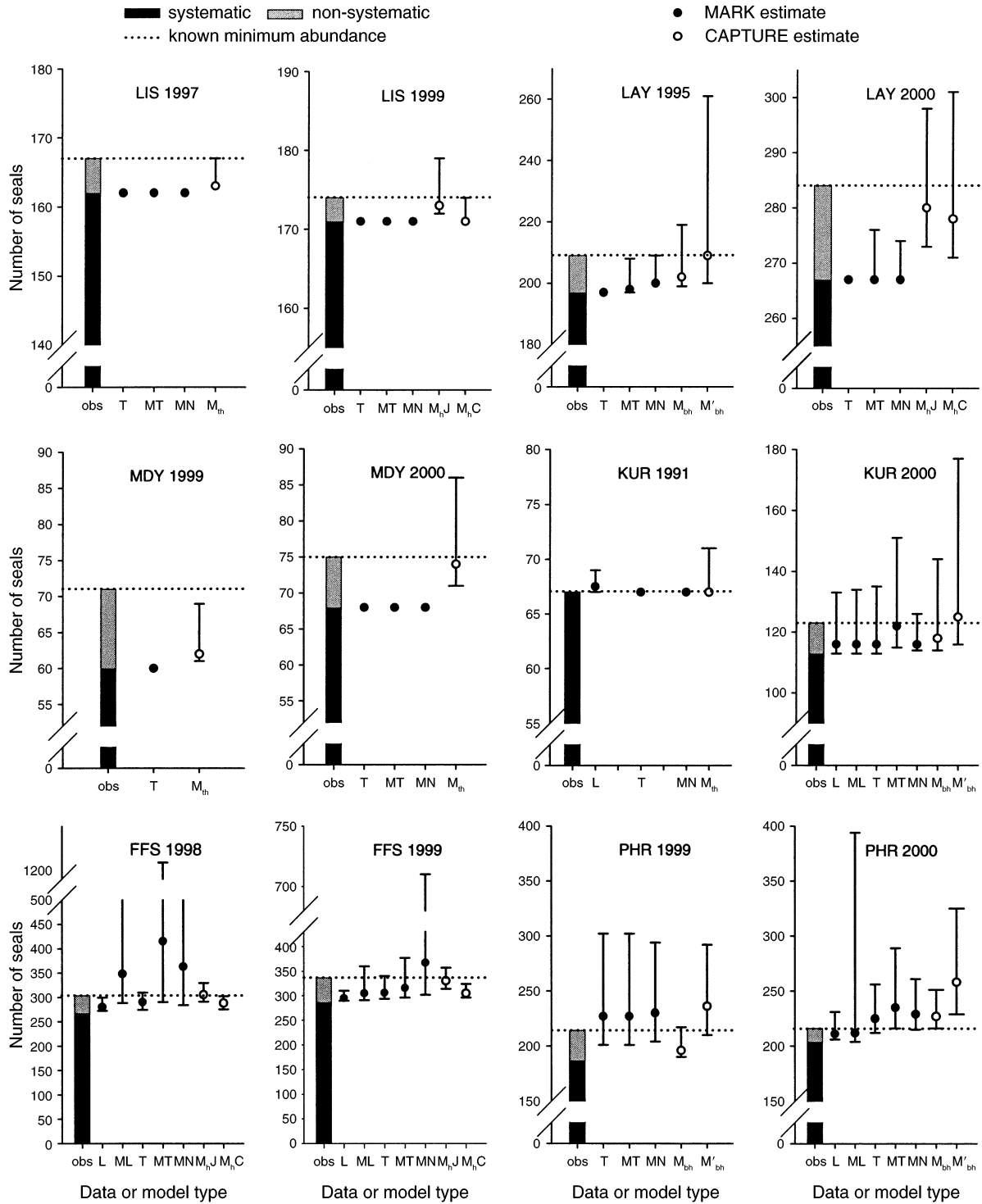


FIG. 3. Comparison of capture-recapture estimates with known minimum population abundance of Hawaiian monk seals for 12 data sets (sites are as in Fig. 2). The observed number of animals seen during the census (obs; black bar) plus those seen during noncensus activities (gray bar) constitute the known minimum abundance (dotted line). Estimated abundances from various Program MARK models are indicated by solid circles with 95% confidence intervals. Lack of confidence intervals indicates that the estimated error was essentially zero. Model types are: L, Location (grouped according to islet where first sighted); ML, mixture model with location; T, adults grouped according to tag status; MT, mixture model with tag status; MN, null mixture model. Open circles and 95% confidence intervals indicate Program CAPTURE estimates. Model types are coded:  $M_{nJ}$ , jackknife estimator with individual heterogeneity;  $M_{nC}$ , Chao's heterogeneity estimator;  $M_{nh}$ , time dependence and heterogeneity;  $M_{bh}$ , removal method behavioral response and heterogeneity; and  $M'_{bh}$ , Pollock and Otto behavior response and heterogeneity.

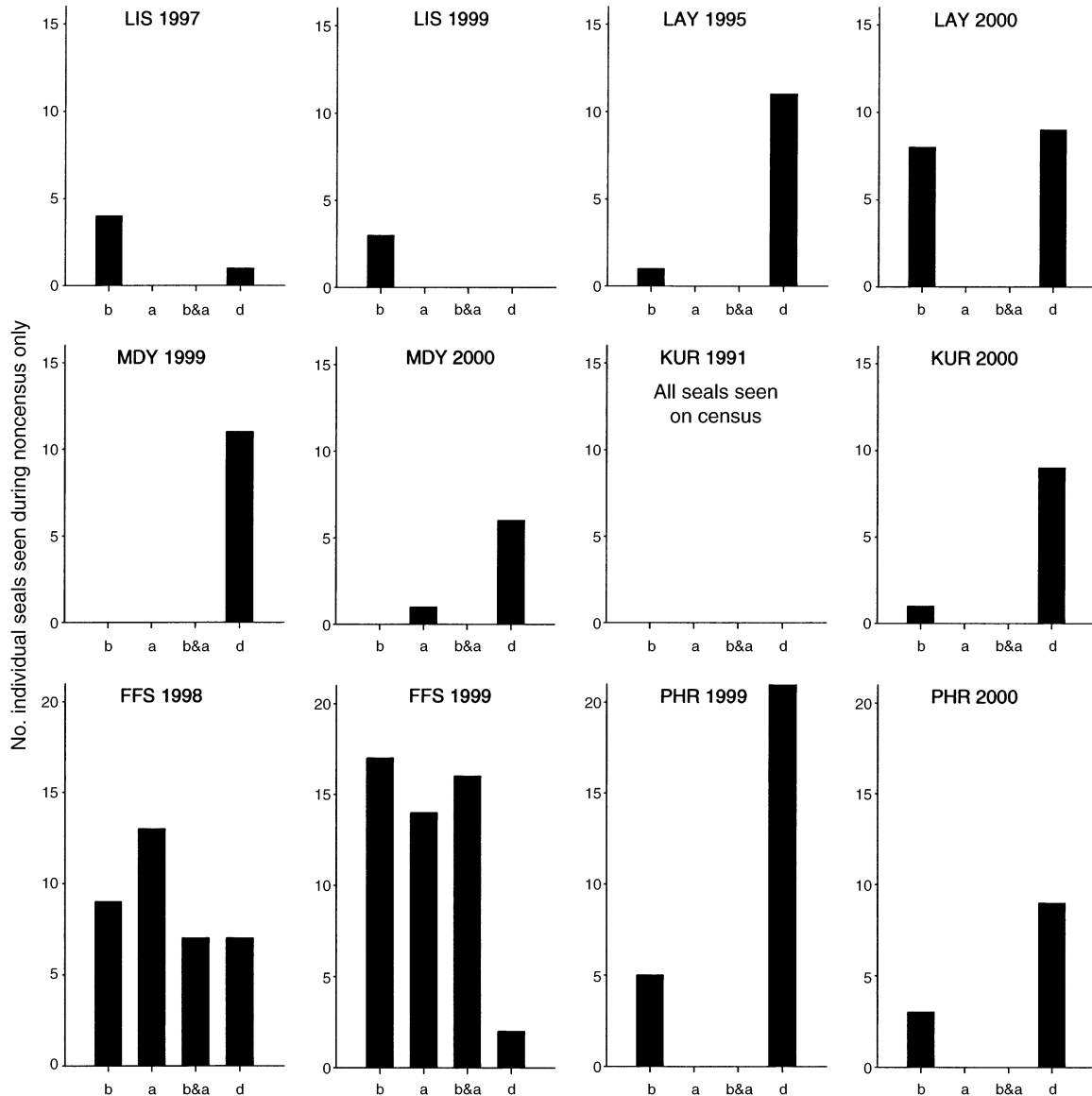


FIG. 4. For 12 modeled data sets (sites are as in Fig. 2), the number of individuals seen only during noncensus activities is classified according to the timing of their sightings relative to the census period: b, only before; a, only after; b&a, only before and after, but not during; and d, during the census period but not during any actual census survey.

No individual characteristics were found to consistently explain why certain individuals were “trap shy” (not sighted during census). Logistic regression analysis of the 12 data sets examined found that adult tag status and islet location had no significant influence on whether seals were seen on systematic surveys. For both years at Lisianski Island and for Kure Atoll in 1991, too few seals were unseen on censuses to allow statistical analysis. At Laysan Island (both years) and Kure Atoll in 2000, no significant effects were detected.

At the remaining sites, age and sex class were sometimes related to whether seals were seen on census, but no consistent pattern emerged. At Midway Atoll, there

was some suggestion that juveniles were more likely to be seen during census than were adults or subadults. All Midway juveniles were sighted on at least one census, whereas 20% of older animals were not seen on any census in 1999 ( $P = 0.06$ ) and 12% in 2000, ( $P = 0.33$ ). In contrast, at Pearl and Hermes Reef, juveniles tended to be less likely seen on censuses than older seals in 1999 ( $P = 0.06$ ). In 2000, juvenile females were less likely to be seen than all other groups ( $P = 0.004$ ). Finally, at French Frigate Shoals in 1999, there was a strong tendency for males to be missed on census compared to females, without regard to size class ( $P = 0.008$ ). At the same site in 1998, no significant effects were found.

## DISCUSSION

All data sets examined involved small populations (<500 individuals) with ample capture occasions, and analyses were stratified with respect to characteristics believed likely to influence capture probability (i.e., sex, size, etc.). It is notable that even under these seemingly favorable conditions, the capture–recapture abundance estimates obtained were quite often poor, mainly due to negative bias. Eberhardt et al. (1999) applied a relatively simple capture–recapture method (Manly–Parr; Seber 1982) to Hawaiian monk seal census data from Laysan Island and also obtained estimates that were negatively biased.

Two assumptions, population closure and homogeneity of capture probabilities, were most likely subject to violation. Because young-of-the-year were excluded from the analyses, lack of population closure could involve only death, immigration, or emigration during the capture period, all of which could have occurred to some extent. However, lack of population closure within the census period will tend to negatively bias capture probability estimates, thereby causing an overestimation of abundance (Otis et al. 1978). Thus, violation of population closure could not account for the pattern of underestimation found in the present analysis. If a significant number of animals had been identified on nonsystematic sightings and then subsequently died prior to the systematic capture period, this could indeed lead to capture–recapture estimates lower than the known minimum abundance. However, the number of animals seen only before the census period could account for only a relatively small portion of the negative bias (Fig. 4).

In contrast, individual capture heterogeneity leads to negative bias in population estimates (Otis et al. 1978). It is clear from the significant effects of sex, size, tag status, and location (Table 1) that all individual monk seals did not have equal capture probabilities. Yet estimates remained negatively biased even after stratification, indicating that individual capture heterogeneity probably existed that could not be ascribed to these characteristics. It is possible that the particular habits of seals coming and going to land and sea, individual daily activity patterns, or other factors yielded individual capture heterogeneity.

In some cases, mixture models (Norris and Pollock 1995, Pledger 1998, 2000) yielded upper confidence limits that included the known minimum abundance (Fig. 3). However, the reduction in bias came at a high cost in terms of precision. Estimates from most of the mixture models were so uncertain that they would be virtually useless for monitoring trends over time.

Program CAPTURE performed best among the model types examined and, although still prone to negative bias, it may nevertheless prove useful in characterizing population trends in Hawaiian monk seals. Notably, the model selection algorithm always gave highest rank to

estimators involving capture heterogeneity (sometimes with time dependence and/or behavioral response). Thus, varying combinations of factors seemed to influence the capture histories observed in different years at various sites, whereas individual heterogeneity was always present.

Because they partition their time between land and sea (and typically can only be captured on land), pinnipeds may exhibit especially high levels of capture heterogeneity compared to either entirely terrestrial or aquatic animals. Nevertheless, other studies that appraised capture–recapture estimates relative to known abundance suggest that heterogeneity is common and that failing to validate model performance or assumptions may lead to incorrect inference. Even in a very controlled experimental situation, Edwards and Eberhardt (1967) found that capture–recapture estimates of a known number of cottontails in an enclosure were negatively biased. Pledger (2000) reanalyzed Edwards and Eberhardt's cottontail data using a two-mixture heterogeneity model and found that although estimated  $N$  was close to the known value, the upper confidence limit was very high.

Although it is rare relative to the overall application of capture–recapture methods, researchers have found various clever ways to evaluate abundance estimators for their particular species of interest under natural conditions (Strandgaard 1967, Mares et al. 1981, Greenwood et al. 1985, Boulanger and Krebs 1994, Manning et al. 1995, Koper and Brooks 1998, Cross and Waser 2000). All of these studies save one (Cross and Waser 2000) found evidence for capture heterogeneity that led to negative bias among estimators. The studies that evaluated Program CAPTURE, along with other methods, found that CAPTURE performed best (particularly the heterogeneity models:  $M_{th}$ ,  $M_{bh}$ ,  $M_{th}$ ). This pattern is consistent with the monk seal results. Authors often, but not always (Koper and Brooks 1998), concluded that they had identified an estimator that, although to some degree was biased, was adequate for their purposes. Universally, it was clear that, without independent validation, there would have been little basis for model selection.

The Hawaiian monk seal data are quite unique in that it has been possible to obtain good minimum abundance estimates for some subpopulations and years. Without this information, one would have been tempted to erroneously conclude that various estimates were reliable. Although such reference points are rarely available, a practical general approach (used here and in the validation studies previously described), is to select representative, tractable demographic units for which simultaneous capture–recapture sampling and exhaustive enumeration can be accomplished. The model or models that perform best can subsequently be applied to the broader population of interest with greater confidence and justification. Dorazio and Royle (2003) similarly suggest that prior knowledge about

sources of heterogeneity is desirable to help select among classes of mixture models. This seems especially germane when one is keenly interested in why a certain model is best. In the practice of conservation and management, it frequently may be that determining which model type produces the most reliable inference is paramount, regardless of the underlying dynamics. In such circumstances, the approach offered here may be helpful.

New capture-recapture methods are continually developed and added to a long list of abundance estimators designed to better address heterogeneity and other issues (e.g., Schwarz and Seber 1999, Yip et al. 2000, Durban 2002, Dorazio and Royle 2003). The capture-recapture literature is replete with numerous simulations and fewer empirical studies confirming that valid inference can be quite sensitive to model selection. Yet, in practice, the choice is all too often made in the absence of corroborating information.

#### ACKNOWLEDGMENTS

I thank Jeff Laake, whose consultation at every stage of this project was absolutely invaluable. John Durban and Vladimir Groisbois provided helpful insight regarding modeling approaches and assistance with Program MARK. Thea Johanos' intimate knowledge of the monk seal database and data collection protocols greatly improved the analyses. Don Bowen, Lee Eberhardt, and Don Siniiff deserve recognition for their encouragement to explore the applicability of capture-recapture methods to the monk seal data. I am grateful to Gerard Dinardo and an anonymous reviewer of this paper. Ani Au's assistance in obtaining relevant literature was greatly appreciated. Hawaii Wildlife Fund provided data for Midway Atoll in 1999. Finally, thanks to Bud Antonelis and Paul Thompson for their advice, support, and comments on the manuscript.

#### LITERATURE CITED

- Anderson, D. R., K. P. Burnham, and W. L. Thompson. 2000. Null hypothesis testing: problems, prevalence, and an alternative. *Journal of Wildlife Management* **64**:912-923.
- Boulanger, J., and C. J. Krebs. 1994. Comparison of capture-recapture estimators of snowshoe hare populations. *Canadian Journal of Zoology* **72**:1800-1807.
- Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference: a practical information-theoretic approach. Springer-Verlag, New York, New York, USA.
- Burnham, K. P., and W. S. Overton. 1978. Estimation of the size of a closed population when capture probabilities vary among animals. *Biometrika* **65**:625-633.
- Burnham, K. P., and W. S. Overton. 1979. Robust estimation of population size when capture probabilities vary among animals. *Ecology* **60**:927-936.
- Carretta, J. V., M. M. Muto, J. Barlow, J. Baker, K. A. Forney, and M. Lowry. 2002. U.S. Pacific Marine Mammal Stock Assessments: 2002. NOAA, La Jolla, California, USA. **NOAA-TM-NMFS-SWFSC-346**.
- Chao, A. 1989. Estimating population size for sparse data in capture-recapture experiments. *Biometrics* **45**:427-438.
- Chao, A., S.-M. Lee, and S.-L. Jeng. 1992. Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal. *Biometrics* **48**:201-216.
- Craig, M. P., and T. J. Ragen. 1999. Body size, survival, and decline of juvenile Hawaiian monk seals, *Monachus schauinslandi*. *Marine Mammal Science* **15**:786-809.
- Cross, C. L., and P. M. Waser. 2000. Estimating population size in the banner-tailed kangaroo rat. *Southwestern Naturalist* **45**:176-183.
- Dorazio, R. M., and J. A. Royle. 2003. Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics* **59**:351-364.
- Durban, J. W. 2002. Bayesian methods for marine mammal population assessment. Dissertation. University of Aberdeen, Aberdeen, Scotland, UK.
- Eberhardt, L. L., R. A. Garrott, and B. L. Becker. 1999. Using trend indices for endangered species. *Marine Mammal Science* **15**:766-785.
- Edwards, W. R., and L. L. Eberhardt. 1967. Estimating cottontail abundance from live trapping data. *Journal of Wildlife Management* **31**:87-96.
- Greenwood, R. J., A. B. Sargeant, and D. H. Johnson. 1985. Evaluation of mark-recapture for estimating striped skunk abundance. *Journal of Wildlife Management* **49**:332-340.
- Harting, A. L. 2002. Stochastic simulation model for the Hawaiian monk seal. Dissertation. Montana State University, Bozeman, Montana, USA.
- Johanos, T. C., B. L. Becker, and T. J. Ragen. 1994. Annual reproductive cycle of the female Hawaiian monk seal (*Monachus schauinslandi*). *Marine Mammal Science* **10**:13-30.
- Kendall, W. L., 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.
- Kenyon, K. W., and D. W. Rice. 1959. Life history of the Hawaiian monk seal. *Pacific Science* **13**:215-252.
- Koper, N., and R. J. Brooks. 1998. Population-size estimators and unequal catchability in painted turtles. *Canadian Journal of Zoology* **76**:458-465.
- Manning, T., W. D. Edge, and J. O. Wolff. 1995. Evaluating population-size estimators: an empirical approach. *Journal of Mammalogy* **76**:1149-1158.
- Mares, M. A., K. E. Streilein, and M. R. Willig. 1981. Experimental assessment of several population estimation techniques on an introduced population of eastern chipmunks. *Journal of Mammalogy* **62**:315-328.
- Norris, J. L., and K. H. Pollock. 1995. A capture-recapture model with heterogeneity and behavioural response. *Environmental and Ecological Statistics* **2**:305-313.
- 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**:1-135.
- Pledger, S. 2000. Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics* **56**:434-442.
- Pledger, S. A. 1998. Finite mixtures in closed capture-recapture models. Research Report 98-18, School of Mathematical and Computing Sciences, Victoria University of Wellington, Wellington, New Zealand.
- Pollock, K. H., and M. C. Otto. 1983. Robust estimation of population size in closed animal populations from capture-recapture experiments. *Biometrics* **39**:1035-1050.
- Ragen, T. J., and D. M. Lavigne. 1999. The Hawaiian monk seal: biology of an endangered species. Pages 224-245 in J. R. Twiss, Jr., and R. R. Reeves, editors. Conservation and management of marine mammals. Smithsonian Institution Press, Washington, D.C., USA.
- Rexstad, E. A., and K. P. Burnham. 1991. User's manual for interactive Program CAPTURE. Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, Fort Collins, Colorado, USA.
- Schwarz, C. J., and G. A. F. Seber. 1999. Estimating animal abundance: review III. *Statistical Science* **14**:427-456.
- Seber, G. A. F. 1982. The estimation of animal abundance and related parameters. MacMillan, New York, New York, USA.
- Stone, H. S. 1984. Hawaiian monk seal population research, Lisianski Island, 1982. U.S. Department of Commerce, La

- Jolla, California, USA. NOAA Technical Memorandum **NOAA-TM-NMFS-SWFSC-47**.
- Strandgaard, H. 1967. Reliability of the Peterson method tested on a roe-deer population. *Journal of Wildlife Management* **31**:643–651.
- U.S. Department of Commerce. 1976*a*. Hawaiian monk seal proposed “endangered” status. *Federal Register* **41**:33,922–33,924.
- U.S. Department of Commerce. 1976*b*. Hawaiian monk seal final regulations. *Federal Register* **41**:51,611–51,612.
- Van Toorenburg, R. A., W. G. Gilmartin, and J. R. Henderson. 1993. Composition of the Hawaiian monk seal population at Kure Atoll, 1990. *Pacific Science* **47**:211–214.
- White, G. C. 2002. Program MARK. Updated version, 16 December 2002. [Online: (<http://www.cnr.colostate.edu/~gwhite/mark/mark.htm>).]
- White, G. C., D. R. Anderson, K. P. Burnham, and L. Otis. 1982. Capture–recapture and removal methods for sampling closed populations. Los Alamos National Laboratory, Los Alamos, New Mexico, USA.
- Yip, P. S. F., L. Xi, A. Chao, and W.-H. Hwang. 2000. Estimating the population size with a behavioral response in capture–recapture experiment. *Environmental and Ecological Statistics* **7**:405–414.
- Zar, J. H. 1984. *Biostatistical analysis*. Second edition. Prentice-Hall, Englewood Cliffs, New Jersey, USA.