Heterogeneity and the mark–recapture assessment of the Scotian Shelf population of northern bottlenose whales (*Hyperoodon ampullatus*)

Hal Whitehead and Tonya Wimmer

Abstract: A population of northern bottlenose whales (*Hyperoodon ampullatus*) inhabits the waters along the edge of the Scotian Shelf. The most important habitat of this population is the Gully, a large submarine canyon, where animals were photographically identified between 1988 and 2003. Open mark–recapture models, including mixture models that allow for heterogeneity in identifiability and (or) mortality among individuals, were fitted to identification-history data. Models without heterogeneity in identifiability had poor fit to the data and underestimated population size. The population is estimated to contain about 163 animals (95% confidence interval 119–214), with no statistically significant temporal trend. About 12% of the population has a high probability of being identified within the Gully in any year. Many of them are mature males. The remainder is less likely to be identified in the Gully during any year, spend generally shorter periods in the Gully even in years when they are found, and are more likely to be female. This and other work indicate a poorly mixed population inhabiting the canyons and other deeper waters off the Scotian Shelf. Nonparametric bootstrap methods were used to validate the estimation procedure and to estimate the efficiency of future fieldwork scenarios.

Résumé : Une population d'hypéroodons arctiques (*Hyperoodon ampullatus*) vit dans les eaux le long de la plate-forme néo-écossaise. L'habitat le plus important pour cette population est le Gully, un grand canyon sous-marin, où les animaux ont été identifiés par photographies entre 1988 et 2003. Nous avons ajusté aux données des suivis des identifications des modèles ouverts de marquage et de recapture, dont des modèles de mélange qui tiennent compte de l'hétérogénéité individuelle de la fiabilité des signalements et (ou) de la mortalité. Les modèles qui ne tiennent pas compte de l'hétérogénéité des signalements s'ajustent mal au données et sous-estiment la taille de la population. La population contient, selon notre estimation, environ 163 individus (intervalle de confiance de 95 %, 119–214) et ne montre pas de tendance temporelle significative. Environ 12 % des individus de la population ont une forte probabilité d'être signalés dans le Gully durant l'année. Plusieurs d'entre eux sont des mâles matures. Les autres sont moins susceptibles d'être signalés dans le Gully dans une année ou l'autre, ils passent généralement moins de temps dans le Gully même les années où on les y retrouve et ils ont plus de chances d'être des femelles. Notre étude et d'autres indiquent que c'est une population peu intégrée qui habite les canyons et les autres eaux profondes au large de la plate-forme néo-écossaise. Des méthodes de bootstrap non paramétriques ont servi à valider notre processus d'estimation et à évaluer l'efficacité des différents scénarios futurs de recherche sur le terrain.

[Traduit par la Rédaction]

Introduction

Population monitoring is key to both management and conservation. One of the principal techniques used in wildlife monitoring is mark–recapture, in which animals are individually identified in some way, and then the patterns of recapture, or re-identification, are used to infer population parameters, including population size (e.g., Seber 1992). Crucial assumptions of almost all the statistical techniques used in mark–recapture analyses are that all animals have the same probability of surviving between sampling periods and the same probability of being recaptured in any period. Especially with open populations, in which individuals may enter or leave the population between sampling periods, the possibility that there is heterogeneity in either mortality or rate of identification has received rather little attention, although heterogeneity may substantially bias mark–recapture population estimates (e.g., Pollock 1982). Here we investigate the effects of heterogeneity on mark–recapture estimates of population size and population trend for a set of animals that are a focus of conservation attention.

Northern bottlenose whales (*Hyperoodon ampullatus*) are residents of the northern North Atlantic. The southernmost location where they can be found reliably is in the Gully, a submarine canyon on the edge of the Scotian Shelf (44°N, 59°W). Photoidentification studies in the Gully since 1988 have allowed mark–recapture assessments of the size and

Received 8 November 2004. Accepted 20 May 2005. Published on the NRC Research Press Web site at http://cjfas.nrc.ca on 19 October 2005.
J18403

H. Whitehead[1] and T. Wimmer. Department of Biology, Dalhousie University, Halifax, NS B3H 4J1, Canada.

[1] Corresponding author (e-mail: hwhitehe@dal.ca).

doi: 10.1139/F05-178

other characteristics of the Gully population (Whitehead et al. 1997; Gowans et al. 2000*a*). The most recent assessment (Gowans et al. 2000*a*), using data collected between 1988 and 1999, suggested a temporally stable population of about 130 individuals using the Gully, although only about 34% of the population were estimated to be in the canyon at any time. The small size of this population, together with threats from entanglement in fishing gear and the expansion of oil and gas exploration and development in the waters surrounding the Gully, led the Scotian Shelf population of northern bottlenose whales to be assessed as endangered by the Committee on the Status of Endangered Wildlife in Canada in 2002 (COSEWIC 2002). The presence and predicament of northern bottlenose whales in the Gully were also important factors in the development of the Gully Marine Protected Area created in 2004 (http://www.dfo-mpo.gc.ca/media/backgrou/2004/hq-ac61a_e.htm). Thus, from the conservation perspectives of both endangered species and protected areas, it is important to monitor the status of the northern bottlenose whales that inhabit the Gully.

Previous assessments of the population of northern bottlenose whales using the Gully (Whitehead et al. 1997; Gowans et al. 2000*a*) were limited by a lack of information of geographical population structure. Since 2000, studies have improved our knowledge of the distribution and movements of animals outside the Gully. Substantial, and statistically significant, differences have been found between the distributions of mitochondrial haplotypes among animals in the Gully and those from samples of whales killed by Norwegian whalers in the Labrador Sea, the closest area to the Scotian Shelf where bottlenose whales can be found reliably (Dalebout et al. 2001; M. Dalebout, Department of Biology, Dalhousie University, Halifax, NS B3H 4J1, personal communication). There were also no matches between a sample of seven photoidentifications collected off Labrador in 2003 and the Scotian Shelf catalogue. The Labrador Sea concentration is about 2000 km from the Gully through the deeper (>500 m) waters that are the habitat of this species.

We have also looked for bottlenose whales closer to the Gully. A survey of the 1000 m depth contour between New Jersey and the southern Grand Banks in 2001 found northern bottlenose whales in the Gully and in Shortland and Haldimand canyons, which also lie on the edge of the Scotian Shelf, 50 km and 100 km east of the Gully, respectively, but nowhere else (Wimmer and Whitehead 2004). The presence of bottlenose whales in Shortland and Haldimand canyons was confirmed in 2002. While some of the animals photographically identified in Shortland and Haldimand canyons in 2001 and 2002 were known from the Gully, and a number of movements between the canyons were documented, there was incomplete mixing, with some animals showing preferences for particular canyons (Wimmer and Whitehead 2004). This suggests that a population size estimated just from data collected in the Gully (as in Whitehead et al. 1997 and Gowans et al. 2000*a*) may be an underestimate for the entire Scotian Shelf population and that the population being identified in the Gully is heterogeneous in identification rates.

Heterogeneity, in which individuals have substantially different probabilities of survival or identification, is a major source of potential error in mark–recapture studies (e.g., Hammond 1990). The rate of identification, or identifiability, of an animal is here defined as the probability that if only one member of the population is identified, it is that animal. Identifiabilities may vary among members of a population because of differences in behaviour, including preferred group size, variation in the quality of the markings used to make identifications, movement patterns, or habitat use. The results from the photoidentifications in the three canyons showed variation among individuals in movement patterns and habitat use, and thus potential heterogeneity in identifiability within the Gully.

With this new information about geographic population structure, together with additional photoidentifications collected during the period 2000–2003 that were not available for the most recent assessment (Gowans et al. 2000*a*), it made sense to reassess this population, incorporating and examining the issue of heterogeneity in identification rates within the Gully. Thus the primary questions being addressed in this paper are as follows: what is the size of the Scotian Shelf population of northern bottlenose whales; is there a temporal trend in numbers; is there evidence for heterogeneity in identification rates; and, if there is such evidence, what are the characteristics of individuals with high and low rates of identification in the Gully? To address these questions, we use modifications of newly developed, open mark–recapture methods, which incorporate heterogeneity using mixture models (Pledger et al. 2003). As the technique is new, its performance is tested using parametric bootstrap methods: simulated data sets are produced using the assumed population model and estimated parameters and are then reinserted into the estimation models (see Buckland and Garthwaite 1991). Comparisons between the input and output parameters allow bias and precision to be estimated.

Monitoring the health of this population in the future will be important from both the endangered species and protected areas perspectives, yet photoidentification research is costly in both financial terms and, potentially, in terms of disturbance to the animals and the protected area. Thus, we also examine the efficiency of different sampling regimes for monitoring future trends in the population of Scotian Shelf bottlenose whales.

Methods

Fieldwork

Northern bottlenose whales were individually identified using photographs taken in the Gully during the summers between 1988 and 2003, although no data usable for this analysis were collected in 1991, 1992, and 2000, years in which effort was very low (Table 1). Most field work was carried out from 10 m (1988–1990) or 12m (1993–2003) auxiliary sailing vessels, but additional data were collected from larger vessels (during surveys of the US National Marine Fisheries Service and the Department of Fisheries and Oceans Canada) or inflatables deployed from them in 2001, 2002, and 2003. The animals were photographed, usually using 35 mm cameras with 300 mm lenses and black and white film, although digital cameras were used during a few of the later studies. When animals were closer than about 30 m from the vessel, photographs were taken of both the left and right flanks and dorsal fins, as well as the melon (forehead). Photographs were taken whenever animals were

Table 1. Summary of photoidentification data collected in the Gully, 1988–2003.

	No. of reliable, $Q \geq 4$ identification photographs		No. of reliably marked animals identified	
Year	Left side	Right side	Left side	Right side
1988	8	20	4	5
1989	245	203	45	38
1990	315	340	61	62
1993	93	81	16	23
1994	48	48	19	25
1995	5	24	4	12
1996	286	268	49	45
1997	265	218	47	54
1998	386	301	40	37
1999	88	53	25	21
2001	35	36	15	14
2002	50	56	16	16
2003	49	67	12	17

Note: See text (Analyzing photoidentifications in Methods) for description of quality (Q) grading system.

within range and (except in 1999) of all animals in the group irrespective of obvious markings or whether the animal had been previously photographed.

Analyzing photoidentifications

Photoidentifications were processed following the protocol described by Gowans and Whitehead (2001). All images were given a quality (Q) grade ranging from $Q = 1$ (poor) to $Q = 6$ (excellent) based upon the characteristics of the image (focus, exposure, angle of fin relative to the negative plane, and proportion of image filled by fin), but independent of the marks on the fin and surrounding area. Only images with $Q \geq 4$ were used in subsequent analyses as recommended by Gowans and Whitehead (2001).

Some of the marks found on the fins and flanks of northern bottlenose whales change over intervals of years (Gowans and Whitehead 2001). Thus this population analysis was restricted to animals marked with notches on the dorsal fin, indentations on the back, and mottled patches for none of which there was evidence of change with time (Gowans and Whitehead 2001). These are referred to as reliably marked individuals. The proportion of $Q \geq 4$ photographs showing reliably marked individuals and its inverse, the ratio of all photographs to those showing reliably marked animals, were calculated for all years combined (excluding 1999, when photographs were not taken randomly with respect to markings). Coefficients of variation for both these ratios were estimated from the variation in the ratios calculated separately for each of the years, weighting by the number of $Q \geq 4$ photographs taken during the year for the proportion of reliably marked individuals, and the number of reliably marked photographs for its inverse.

As each year's data were analyzed, photographs were matched against the catalogue built up from previous years as described by Gowans and Whitehead (2001). Data from photographs showing the left and right sides of animals were processed separately (except for the calculation of the proportion of reliably marked individuals, as reliable marks in dorsal fins, such as nicks, are visible from both sides; Gowans and Whitehead 2001).

Where possible, animals were assigned to one of three age–sex classes (mature males, subadult males, and females–immatures) using photographs of their melon (forehead) area as described by Gowans et al. (2000*b*).

Population analysis

Assessments of the population of reliably marked individuals using the Gully were made with a variety of population models (including those used by Gowans et al. 2000*a*). To incorporate the possibility of heterogeneity in either mortality or identification rate, we constructed two-component finite mixture models following the methods of Pledger et al. (2003). The models are formally described in Appendix A. They all condition on the first identification of each individual and assume overall identification rates in each year j of n_j/N_j (the number of animals identified during the year divided by the estimated population size that year, as in Seber 1982, p. 557). Heterogeneity is incorporated by having two classes of individual (A and B), which may have different identification and (or) mortality rates. The population can also increase or decrease at a constant rate per year. Thus, the full model has the following parameters: N_M, population of reliably marked animals using the Gully at midpoint of sampling (1995–1996); r, proportional rate of increase in population per year; α, proportion of population in class A (with $1-\alpha$ in class B); β, ratio of identifiability of class B animals compared with class A animals; δ_A, mortality rate of class A animals per year; δ_B, mortality rate of class B animals per year.

Simpler models were investigated by using the following restrictions or combinations of them: $r = 0$, no trend; $\alpha = 1$, no heterogeneity; $\beta = 1$, no heterogeneity in identification; $\delta_A = \delta_B$, no heterogeneity in mortality; $\delta_A = \delta_B = 0$, $\alpha = 1$, and $r = 0$, closed, homogeneous population. The models examined are listed (Table 2).

Population parameters were estimated using maximum likelihood methods, and their fits were compared by Akaike's Information Criterion (AIC), with the lowest AIC indicating the preferred model (Burnham and Anderson 2002). Confidence intervals (CIs) for the parameter estimates were calculated using both likelihood and nonparametric bootstrap methods with 1000 replicates (see Buckland and Garthwaite 1991). The size of the entire study population, N_T, was estimated by multiplying the estimate for reliably marked individuals, N_M, by the estimate of the ratio of all photographs to those showing reliably marked animals, h. CIs for N_T were then estimated from the CI for N_M and the estimated coefficient of variation of h, so that the upper confidence interval (uci) was given by

$$(1) \qquad \mathrm{uci}(N_T) = N_T + 1.96\, N_T \sqrt{\mathrm{CV}(h)^2 + \left\{\frac{[\mathrm{uci}(N_M) - N_M]}{1.96\, N_M}\right\}^2}$$

where CV is the coefficient of variation. The lower CI was calculated similarly.

Goodness-of-fit measures and tests for open-population mixture models of heterogeneity are not well developed

Table 2. Summary of population models fitted to Gully photoidentification data.

Model	Closed? ($\delta_A = \delta_B = 0$, $\alpha = 1$, $r = 0$)	Trend? ($r \neq 0$)	Heterogeneity in: Identifiability? ($\beta \neq 1$)	Heterogeneity in: Mortality? ($\delta_A \neq \delta_B$)	AIC: Left	AIC: Right
Closed (Schnabel)	Yes	No	No	No	980.69	985.53
Mortality	No	No	No	No	870.69	882.00
Mortality + trend	No	Yes	No	No	871.59	883.99
Heterogeneity in mortality	No	No	No	Yes	861.54	872.73
Heterogeneity in mortality + trend	No	Yes	No	Yes	862.46	874.21
Heterogeneity in identifiability	No	No	Yes	No	839.56	843.56
Heterogeneity in identifiability + trend	No	Yes	Yes	No	841.19	845.11
Heterogeneity in mortality & identifiability	No	No	Yes	Yes	836.12*	843.38
Full: heterogeneity in mortality & identifiability + trend	No	Yes	Yes	Yes	836.45	843.31*

Note: Akaike's Information Criterion (AIC) is given for the data from the left and right side photographs.
*The lowest AIC indicates the best fit to the data for each side.

(Pledger et al. 2003), and some of the more standard tests used with nonheterogeneous models (e.g., TEST 3 of Burnham et al. 1987) are clearly inappropriate. Here we compare the observed and expected distributions of the number of years during which individuals were identified.

To examine the performance of the mixture-model estimation procedure, we used parametric bootstraps, constructing artificial data sets using the identification schedules and rates of the real data and parameter estimates (see Appendix B), and then running them through the estimation procedure. One thousand data sets were constructed using each of the "heterogeneity in identifiability" and "heterogeneity in mortality & identifiability" models, as well as using a model with continuous variability in identifiability among individuals (see Appendix B). Each data set was tested using several of the models, with that producing the lowest AIC being noted.

Using Bayes' theorem, individuals were assigned posterior probabilities to classes A and B (Pledger et al. 2003) by the "heterogeneity in mortality & identifiability" model using the left-side photographs. The following characteristics of individuals were then compared with these posterior probabilities to look for traits of individuals with high and low identification rates: identification histories within and between years, age–sex class, prevalence of markings, and mean group size (with group being defined as whales within five body lengths of one another and showing coordinated behaviour (Gowans et al. 2001); group sizes were recorded when photographs were taken; only one group size was used for each individual in each 15 min interval; only those group size records made with left-side photographs were used). These analyses used the SOCPROG programs, and modifications of them, written in MATLAB (see http://myweb.dal.ca/~hwhitehe/social.htm).

Planning future effort

The measure of effort used was days spent at sea per year. This varied between 4 and 69 days·year^{-1}. The number of reliably marked animals identified using left or right sides during the year j (n_j) was plotted against effort in that year (f_j). A logistic model with parameter π was fitted through the data (data for years 2001–2003 were omitted, as significant effort was spent outside the Gully in these years) using a generalized linear model. Thus, the expected number of identifications in year j, $E(n_j)$, was related to population size (N_M, estimated in the earlier part of the paper) by

$$(2) \qquad E(n_j) = \frac{N_M \pi f_j}{1 + \pi f_j}$$

Using this relationship, the number of animals identified during a given year could be estimated given the effort expended. However, error about this line was found to be greater than predicted by a binomial model, so in simulating data, a normal approximation was used:

$$(3) \qquad n_j = \text{Normal}\left(\text{Mean} = \frac{N_M \pi f_j}{1 + \pi f_j}, \text{SD} = d\,\frac{\sqrt{N_M \pi f_j}}{1 + \pi f_j}\right)$$

The dispersion parameter, d, was chosen so that the standard deviations of the residuals were similar in the real and simulated data. Simulated values of n_j were rounded to the nearest integer and bounded by 0 and N_M. We simulated 1000 data sets using the parametric bootstrap method (see above; Appendix B) over a 20 year period. For each of these data sets, for five total levels of field effort during the 20 years (100, 200, 400, and 800 days) and for four patterns of effort (equal amount of effort every year, every 2 years, every 5 years, and every 10 years), we used the relationship in eq. 3 to give the number of animals identified in each year of effort and then used the procedures described in Appendix B to produce identification histories. For each data set, each level of identification effort, and each pattern of identification, the simulated data set was run through the full estimation model. The dispersions of estimated trends and population sizes were taken as an indication of the utility of the effort scenario.

Results

Model fitting

The results of fitting the population models are summarized (Table 2). While the full model showed the best fit (lowest AIC) for the right-side photographs, the model without a trend in population size fit best for the left-side photographs, although in each case the differences between the AICs of these two models were small (Table 2). For the

Table 3. Observed and expected distribution of numbers of animals identified during different numbers of years for best model without heterogeneity (Mortality) and full model with heterogeneity in mortality and identifiability, together with results of χ^2 goodness-of-fit tests.

	Left side			Right side		
		Expected under:			Expected under:	
No. of years identified	Observed	Mortality	Full	Observed	Mortality	Full
1	81	64.8	75.7	84	68.7	78.9
2	35	45.0	42.5	39	43.9	42.4
3	17	25.0	19	15	25.2	19.7
4	10	14.2	9.1	9	14.8	9.3
5	6	7.0	4.7	5	7.4	4.6
≥6	11	3.9	9	12	4.1	9.1
χ^2	—	23.3	2.8	—	26.1	2.7
P	—	0.000	0.728	—	0	0.738

right-side photographs, the model that only included heterogeneity in identifiability also had an AIC close to (within 1.0 of) the minimum, indicating substantial empirical support (Burnham and Anderson 2002). However, AICs for all models without heterogeneity in identifiability were all much larger (>15) than the minima (Table 2), indicating that heterogeneity in identifiability is an important feature of these data, whereas heterogeneity in mortality and population trends have marginal significance (see also results of parametric bootstrap analysis, below).

The distribution of the number of years that each animal was observed, together with the expected distributions for the best-fitting model without heterogeneity (the "Mortality" model (the best of those fitted by Gowans et al. 2000*a*)), and the full model with heterogeneity are provided (Table 3). There are considerably more individuals identified in 1 year or during 6 or more years and fewer individuals identified in a total of 2–4 years than expected using models without heterogeneity. This pattern, characteristic of heterogeneous data (Cormack 1985), largely disappears when the model including heterogeneity is used (Table 3), and, at least in this feature, this model fits the data satisfactorily.

Parameter estimates

Parameter estimates for the three best-fitting models, as well as the homogeneous Mortality model, together with their estimated CIs, are shown (Table 4). There is consistency both between estimates from photographs from the left and right sides and between CIs calculated using likelihood and nonparametric bootstrap methods (as well as from the parametric bootstrap analysis; see below, Table 5). The population is estimated to consist of about 115 reliably marked individuals, with a 95% CI of about 95–130. Of the $Q \geq 4$ photographs, 68% (standard error, SE = 3%) were of reliably marked individuals, and the ratio of all photographs to reliably marked photographs was 1.462 (CV = 0.109). By combining this ratio with the population estimates of reliably marked individuals and CIs from nonparametric bootstrap analysis (Table 4) for the best-fitting models (as indicated by the minimum AIC in Table 2) using eq. 1, we calculated total population estimates and CIs for the Scotian Shelf population of northern bottlenose whales. These estimates were 155 animals (95% CI = 114–201 animals) for the left-side photographs and 171 animals (95% CI = 123–227 animals) for the right-side photographs. After averaging, we estimated the population to contain about 163 animals (95% CI = 119–214 animals).

The models that include a trend in population size as well as heterogeneity in identification rates have very similar fits (indicated by AICs) to those that do not include trends. For the left-side photographs, a model without a population trend has the best fit, and for the right-side photographs a model with a trend fits best (Table 2). The best estimates of the trend indicate an increase in population size of about $+2.5\%\cdot\text{year}^{-1}$, but in no case is it statistically significantly different from zero (Table 4).

The population is indicated as being composed of two classes of reliably marked individuals: about 10–25 individuals that are regularly identified in the Gully and have low ($0\% - {\sim}6.5\%\cdot\text{year}^{-1}$) rates of mortality, emigration, or mark change; and about 100 individuals that have roughly 75%–94% lower rate of identification and a higher (${\sim}13\%\cdot\text{year}^{-1}$) rate of mortality, emigration, or mark change. The division into two classes was made for the convenience of fitting the population model, and the results are also consistent with a population containing a continuum of identifiabilities ranging from individuals with high identifiabilities in the Gully to those only likely to be only occasionally identified there (Pledger et al. 2003; see parametric bootstrap analysis below).

Validity of analysis

The validity of these results was checked using the parametric bootstrap analyses (results summarized in Table 5). These analyses indicate that using the minimum AIC criterion, the analyses generally but not always (about 80%–85% of the time) selected the correct input model. Models including trends were sometimes selected, even though there was no trend in the underlying data. Parameter estimates were close to being unbiased (with the exception of N_M and α when the data were produced by the heterogeneity in identifiability models, which were both biased downwards by about 10%). Parametric bootstrap estimates of precision are very similar to those from the nonparametric bootstrap or likelihood methods (compare 95% CIs in Table 4 for left-side data with those from the correct model in Table 5). The

Table 4. Estimates of parameters of those population models with lowest Akaike's Information Criteria (AIC) values, plus the Mortality model, from maximum likelihood.

						Mortality rate (%·year^{-1})	
Data set	Model	Population estimate of reliably marked individuals (N_M)	Trend (%·year^{-1}; r)	Percentage of class A (%; α)	Ratio of identifiability (Class A/Class B; β)	Class A (δ_A)	Class B (δ_B)
Left side	Mortality	86	—	—	—	13.2	
		(77–98)				(9.9–16.9)	
		[76–99]				[10.1–16.8]	
	Heterogeneity in identifiability	107	—	76	0.11	9.1	
		(89–135)		(53–91)	(0.06–0.19)	(5.2–13.3)	
		[91–135]		[45–92]	[0.04–0.15]	[5.4–12.9]	
	Heterogeneity in mortality & identifiability	106	—	90	0.15	14.4	0.0
		(90–128)		(80–96)	(0.08–0.27)	(9.7–20.4)	(0–8.7)
		[90–128]		[60–96]	[0.05–0.26]	[5.2–21.4]	[0–9.6]
	Full: heterogeneity in mortality & identifiability + trend	112	+2.5	89	0.13	13.1	0.0
		(93–139)	(–1.4; +6.4)	(79–95)	(0.07–0.25)	(7.9–19.4)	(0–7.0)
		[91–135]	[–1.6; +6.4]	[70–95]	[0.06–0.30]	[6.9–22.2]	[0–6.6]
Right side	Mortality	89	—	—	—	12.5	
		(79–102)				(9.3–16.2)	
		[78–106]				[9.0–16.3]	
	Heterogeneity in identifiability	112	—	78	0.10	8.6	
		(93–140)		(62–90)	(0.05–0.17)	(5.0–12.5)	
		[95–137]		[58–93]	[0.04–0.15]	[5.0–12.1]	
	Heterogeneity in mortality & identifiability	112	—	86	0.12	12.0	3.3
		(94–139)		(67–96)	(0.06–0.23)	(6.0–17.7)	(0–11.4)
		[93–134]		[61–97]	[0.05–0.23]	[4.3–19.5]	[0–10.6]
	Full: heterogeneity in mortality & identifiability + trend	117	0.031	89	0.13	11.6	0
		(96–146)	(–0.7; +6.7)	(79–96)	(0.07–0.23)	(6.8–17.1)	(0–9.2)
		[96–142]	[–1.0; +6.8]	[68–97]	[0.04–0.24]	[5.3–17.9]	[0–7.9]

Note: Estimated 95% confidence intervals of parameters from likelihood methods are given in parentheses; those from nonparametric bootstrap methods are in square brackets.

Table 5. Summary of results of parametric bootstrap analysis.

							Mortality rate (%·year^{-1})	
Input model	Analytical model	Models selected by AIC	Population estimate of reliably marked individuals (N_M)	Trend (%·year^{-1}; r)	Percentage of class A (%; α)	Ratio of identifiability (Class A/Class B; β)	Class A (δ_A)	Class B (δ_B)
Heterogeneity in identifiability	Input parameters	—	107	—	76	0.11	9.1	
	Mortality	1	84 [75–95]	—	—	—	12.1 [9.0–15.7]	
	Heterogeneity in identifiability	836	98 [85–114]	—	68 [47–84]	0.10 [0.05–0.18]	8.9 [5.9–12.8]	
	Heterogeneity in mortality & identifiability	114	98 [84–114]	—	70 [44–90]	0.12 [0.05–0.21]	9.6 [2.0–17.2]	7.8 [0.1–13.7]
	Heterogeneity in mortality & identifiability + trend	49	99 [83–118]	+0.7 [–2.7; +3.8]	71 [43–92]	0.14 [0.05–0.22]	9.7 [2.0–17.1]	6.9 [0–14.1]
Heterogeneity in mortality & identifiability	Input parameters	—	106	—	90	0.15	14.4	0.0
	Mortality	0	90 [80–101]	—	—	—	12.4 [10.3–17.1]	
	Heterogeneity in identifiability	43	113 [96–133]	—	78 [55–91]	0.11 [0.05–0.18]	9.6 [6.4–12.7]	
	Heterogeneity in mortality & identifiability	816	112 [96–128]	—	89 [79–94]	0.15 [0.07–0.27]	14.6 [9.7–20.6]	0.6 [0.0–5.1]
	Heterogeneity in mortality & identifiability + trend	141	113 [97–133]	+1.1 [–0.7; +4.2]	89 [80–94]	0.15 [0.07–0.26]	14.3 [9.3–20.3]	0.3 [0.0–4.0]
Continuous variability in identifiability	Input parameters	—	107	—	—	—	9.1	
	Mortality	19	86 [76–97]	—	—	—	11.6 [8.8–14.8]	
	Heterogeneity in identifiability	802	96 [82–111]	—	74 [38–96]	0.12 [0.03–0.21]	9.4 [6.4–12.7]	
	Heterogeneity in mortality & identifiability	142	95 [82–111]	—	76 [37–97]	0.37 [0.03–0.42]	10.2 [2.1–15.9]	7.3 [0–14.0]
	Heterogeneity in mortality & identifiability + trend	37	96 [81–117]	0.2 [–3.0; +3.5]	77 [35–98]	0.33 [0.03–0.37]	10.3 [2.0–16.2]	6.6 [0.0–15.8]

Note: For each of three different input models, 1000 data sets were constructed and analyzed using four analytical models. The number of times (out of 1000) that each analytical model was selected using the lowest Akaike's Information Criteria (AIC) is given together with the mean of the estimated parameter values and 95% confidence intervals in brackets.

Table 6. Characteristics of reliably marked animals with posterior probabilities >0.75 of being allocated into either class A (low identification rate and high mortality rate) or class B (high identification rate and low mortality rate).

	Class A	Class B	Tests
No. of animals allocated	143	11	—
Mean years identified	1.7 [1; 5]	7.5 [6; 10]	—
No. of days identified per year (mean difference from annual mean)*	–0.09 [–1.52; 3.45]	0.91 [–0.65; 2.54]	K–W: $P = 0.003$
Proportion of years identified ≥10 days*	0.20 [0; 1]	0.40 [0; 0.8]	K–W: $P = 0.005$
With fin notches (%)	72	91	G: $P = 0.132$
With back indents (%)	9	27	G: $P = 0.099$
With mottling (%)	31	36	G: $P = 0.739$
Mature and subadult males (of sexed animals) (%)	37 ($n = 78$)	64 ($n = 11$)	G: $P = 0.082$
Mature males (of all sexed males) (%)	54 ($n = 26$)	86 ($n = 7$)	G: $P = 0.106$
Mean group size	4.40	4.68	K–W: $P = 0.088$

Note: Patterns of identification between and within years, markings, and sex–age class are indicated. Differences between classes in continuous variables were tested using the Kruskal–Wallis (K–W) test and between categorical variables using likelihood-ratio *G* tests. Most continuous variables ranges are given in square brackets.

*Just using years with substantial identification effort (1989, 1990, 1996, 1997, 1998).

mixture model estimation procedure gives reasonable estimates of population size (downward bias of about 10%) and mortality rate (approximately unbiased) when identifiability is continuous and not bimodal as assumed by the mixture models.

Characteristics of animals with high and low identifiability

Characteristics of reliably marked animals with posterior probabilities >0.75 of being allocated into either class A (low identification rate and high mortality rate) or class B (high identification rate and low mortality rate) are summarized (Table 6). As might be expected (as these were the data used to form the classes), class A animals were identified in fewer years (1–5 years) than class B animals (6–10 years). However, class A animals were also generally identified over shorter time scales within years. Compared with class A animals identified the same year, class B animals were identified on a mean of one additional day and were about twice as likely to be identified over spans of at least 10 days within a year. This indicates a greater short-term identifiability of class B animals. Class B animals were more likely to possess each of the three types of marks that persist over periods of years (notches, back indents, and mottling) and are used to distinguish "reliable" individuals (Gowans and Whitehead 2001), but for none of the mark types was the difference in markings between the classes statistically significant. The majority of class A animals were females, whereas those in class B were more likely to be males, and in particular mature males (Table 6). There was no substantial or statistically significant (at $P < 0.05$) difference between mean group sizes of animals allocated to the two classes (Table 6).

Planning future effort

The relationship between level of effort (days at sea per year) and number of animals identified is shown (Fig. 1). The logistic parameter (π in eq. 2) was estimated to be 0.0113, and the dispersion factor (d in eq. 3) was estimated to be 2.90. The data simulated using eq. 3 appear to have a similar distribution to the real data (Fig. 1). The results of the simulation study on the relationship between effort and the precision of estimated parameters over a 20 year period are shown (Fig. 2). As expected, greater effort gave greater precision. More precise estimates were obtained by spreading effort over annual or biannual studies rather than concentrating time at sea every 5–10 years. The results suggest that to obtain a standard error in trend of less than $1\%\cdot\text{year}^{-1}$ after 20 years requires annual studies of about 20 days each or biannual studies lasting about 40 days.

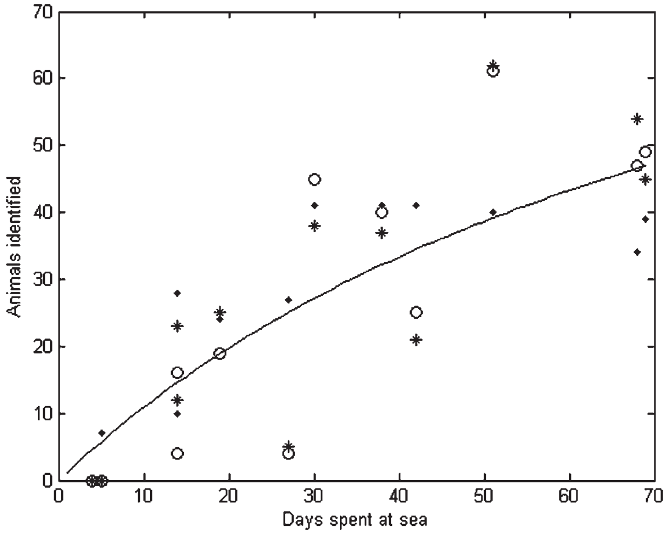
Discussion

Technical issues

Using much of the same data set that our analysis has shown to be clearly heterogeneous in identifiability, Gowans et al. (2000*a*) failed to detect heterogeneity by plotting the residual differences between the observed and expected number of animals with a particular identification history against the number of years identified (as recommended by Cormack 1985). This may be explained by a fairly large number of possible identification histories when there are many sampling periods; then the scatter among the residuals of histories with a particular number of periods identified swamps any bias in the residuals away from zero. Instead, we recommend comparing the observed and expected distributions of the number of periods identified where a poor fit to a model without heterogeneity is clearly indicated.

It is generally considered that heterogeneity in capture rates introduces a downward bias in mark–recapture population estimates (Hammond 1986), but that estimates of survival or mortality are less affected (Pledger et al. 2003). In the case of the real and simulated data sets considered in this paper, the estimates of population size using estimation models without heterogeneity were about 20% lower than those of the mixture models or the population sizes input into the simulations. However, the mortality estimates also appeared biased, upwards by about 30% in the models without heterogeneity compared with those that included heterogeneity in identification or the input mortality in the case of the simulated data sets. These results support the development and use of mixture models to analyse mark–recapture data that contain heterogeneity.

Fig. 1. Relationship between number of days spent at sea in a year and number of reliably marked animals identified in the Gully from left-side photographs (open circles, ○) and right-side photographs (asterisks, *), together with logistic regression curve (solid line), and data simulated using logistic regression curve (solid diamonds, ◆).



A particular case of concern is the quite common use of two-sample Petersen mark–recapture estimates. Sometimes a series of these, comparing identifications in consecutive study periods, is used to investigate trends in population size (e.g., Calambokidis and Barlow 2004). These methods are much simpler to use than the open models considered in this paper and are therefore attractive. However, heterogeneity in identification rates will not only cause such population estimates to be biased (downwards), but the degree of bias will vary with effort. So when such methods were applied to the Gully data set, especially low population sizes were estimated for year pairs such as 1993–1994: 16 (1993) and 19 (1994) left-side reliable identifications with four in common gave a Petersen population estimate of 67 reliably marked individuals. With a population size of 106, 90% of the individuals in class A, and an identifiability ratio of 0.15, the equations in Appendix A give expected numbers identified of 11.1 and 13.5 class A animals in 1993 and 1994, respectively, and 4.9 and 5.5 class B animals in 1993 and 1994, respectively. The expected number identified in both years (ignoring mortality) is then 2.9 for class A and 1.6 for class B, giving a total of 4.5, almost exactly as in the real data. Thus, in populations with heterogeneity in identifiability, Petersen estimates from small sample sizes can substantially underestimate the population size, on account of matches within the sets of more identifiable animals.

In contrast with the poor performance of standard mark–recapture models on heterogeneous data sets, the parametric bootstrap simulation indicated that the two-class mixture models perform quite well on data sets containing continuous variability in identifiability (also noted by Pledger et al. 2003). Thus, a good fit of the two-class model should then not be taken to indicate literally two classes of animal, but rather that there is heterogeneity in the data set, and its extent can be roughly approximated by the standard deviation in identifiability and (or) mortality of the results of the two-class model.

Doubts have been expressed that the use of AIC may overfit mixture models, selecting overly complex models, especially at the boundaries of parameter space (Pledger et al. 2003). In our case, the best-fitting models estimated $\delta_B = 0$, so this was a concern. In the nonparametric bootstrap analysis, the correct model was selected using the lowest AIC criterion in 84% of the runs generated by the heterogeneity in identifiability model and in 82% of the runs generated by the heterogeneity in identifiability and mortality model. More complex models were selected (i.e., overfitting) in 16% and 14% of the runs, respectively, but simpler models (underfitting) occurred in a total of only 44 of 2000 runs. This indicates that while the minimum AIC is most likely to choose the most appropriate model, overfitting may occur quite frequently.

The models used in this paper all assume that the identification rate in each year is the number of identifications in the year divided by the estimated population size that year (as in Seber 1982, p. 557 and Whitehead 1990). This makes the model fitting much more efficient than the standard method in which separate identification probabilities are estimated for each sample (e.g., White and Burnham 1999), as only one parameter is estimated (N_M), but it is not theoretically valid except in the case where the number of identifications collected in each sample was fixed beforehand, which it was not in the Gully. However, the nonparametric bootstrap analyses, which also make this simplifying assumption, indicate that making this assumption introduces no serious bias in parameter estimates or estimates of their precision.

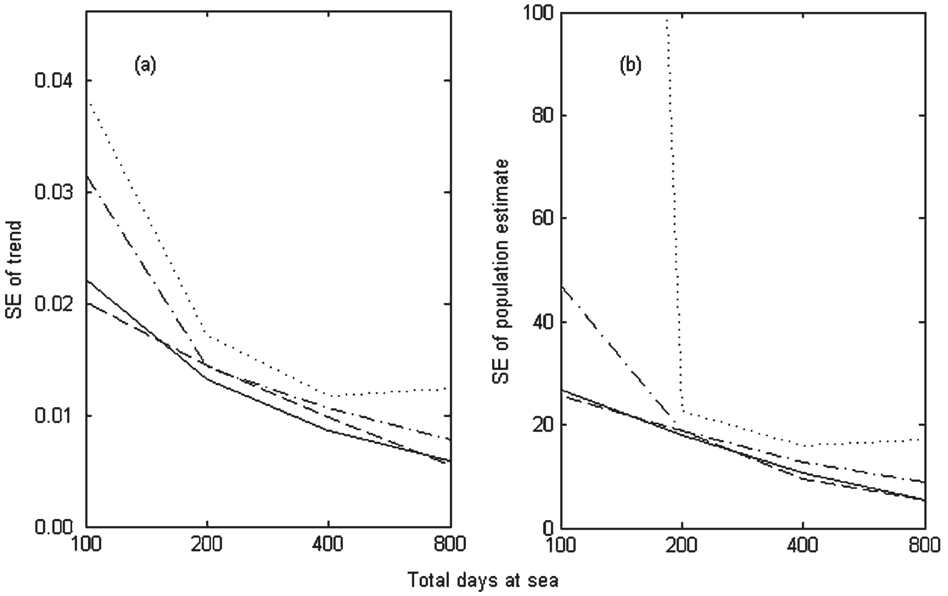
The models that we have used condition on the first identification of each animal. This assumption simplifies the programming, but discards potentially valuable information. In a mark–recapture study of the survival rates of bowhead whales (*Balaena mysticetus*), Zeh et al. (2002) found that estimates were substantially improved when first identification data were used. Adding this to the mixture models that we employed would be challenging, but potentially rewarding.

As a final point in this discussion of technical issues, we would like to emphasize the value of examining the estimation procedure using the parametric bootstrap. Methods of analyzing mark–recapture data using open mixture models are just beginning development (Pledger et al. 2003), and there are no general studies of performance. In such a situation, simulation models can be particularly valuable in supporting the validity of a particular analysis.

Northern bottlenose whales of the Scotian Shelf

This analysis, coupled with the work on the animals outside the Gully (Wimmer and Whitehead 2004), goes some way to filling in the most important gap in our understanding of the population biology of the bottlenose whales that use the Gully: their relationship to other populations of the species. The data collected in 2001 and 2002 and analyzed by Wimmer and Whitehead (2004) showed that there were concentrations of animals in at least two other canyons on the edge of the Scotian Shelf, that animals moved between the different canyons, but that over the 2 year time scale, mixing was incomplete. Some animals, particularly mature males, moved frequently and rapidly between the canyons whereas others, mainly females, showed strong preferences

Fig. 2. The relationship between levels (x axis) and patterns of effort and the precision (indicated by standard error, SE) of estimated (*a*) trend and (*b*) population size from simulation study over a 20 year period. Line types give years between samples: solid line, 1; broken line, 2; dot-dashed line, 5; dotted line, 10.



for particular canyons (Wimmer and Whitehead 2005). Interpreting the output of the population assessments in this context, we suggest that the class B animals, with high identifiability in the Gully and low mortality, consist of many of these mobile mature males, as well as females for whom the Gully is their preferred habitat, whereas the class A animals are principally females whose preferred habitat is outside the Gully but who make infrequent and quite brief visits to the Gully. The characteristics of the class A and class B animals are consistent with this interpretation. However, the apparently higher mortality of the class A animals may well be an artifact. Animals that die during the study, whatever their pattern of usage of the Gully, will tend to be assigned to class A because of their shorter identification history, and animals that use the Gully infrequently may be unidentified during a large part of the end of the study, incorrectly suggesting mortality. As noted above and by Pledger et al. (2003) and illustrated by the parametric bootstrap analysis, the good fit of a model with two classes of identifiability is also consistent with a continuum of identifiabilities among members of the population. We think it likely that this is the case with the Scotian Shelf bottlenose whales and thus that the proportions of animals in the two classes, their relative identifiabilities, and mortalities should not be taken literally.

However, the estimates of population size and trend represent more real traits and are of significance for conservation issues. The population estimates from the best model that did not include heterogeneity, the mortality model, of 86 reliably marked individuals from left-side photographs and 89 from right-side photographs are very similar to those of Gowans et al. (2000*a*): 88 and 84, respectively. This is not surprising, as the analyses used the same data sets with the exception of the 2001–2003 photographs that were not available to Gowans et al. (2000*a*). However, these estimates are about 20% lower than those from assessments using models that include heterogeneity, so that total population estimates (after extrapolation to include the unreliably marked individuals) of 156 and 171 are about 30 animals higher than those from the homogeneous models. This change is entirely due to the more sophisticated estimation procedure, which better accounts for animals that make only occasional use of the Gully. From the arguments of the previous paragraph, the extra animals can be thought of as those whose primary habitat is outside the Gully and rarely visit it. Thus, the new higher estimates are a better reflection of the entire Scotian Shelf population, including animals that are quite resident in the Gully as well as occasional visitors, than previous attempts, although animals that never enter the Gully will still be omitted.

Although the analysis of the right-side photographs showed a marginal preference for a model including a positive trend in population size, the trend was not significantly different from zero, and the parametric bootstrap analysis showed that models including a trend are frequently selected by the minimum AIC criterion even when the population from which the data were collected contains no trend. Thus, the analyses do not support a trend in the Scotian Shelf population size.

Planning future effort

The efficiency of future effort scenarios in detecting trends and estimating population size are underestimated here for at least two reasons. First, we already have 15 years of data, which were not accounted for in the simulations and will form a baseline for continuing work. Second, there is now information on the habitat of the animals outside the Gully. We know that animals can be found in Shortland and Haldimand canyons, and if photographs are taken in these canyons as well as in the Gully (as they were in 2001–2002), this should allow a more complete coverage of the population. With the population more evenly photographed, there will be less heterogeneity in the data set, and the result should be relatively more precise estimates of the remaining population parameters, especially the population size and its temporal trend.

Acknowledgements

Thanks go to all the crew members who collected the data over the years of study. The photoidentifications were analyzed by Annick Faucher, Shannon Gowans, and Jakobina Arch, as well as TW. Our field work and analysis were funded by World Wildlife Fund Canada – Environment Canada Endangered Species Recovery Fund, the Natural Sciences and Engineering Research Council of Canada, the Canadian Whale Institute, the Island Foundation, the Whale and Dolphin Conservation Society, the National Geographic Society, and the Department of Fisheries and Oceans Canada. Photographs from the US National Marine Fisheries Service cruises are courtesy of Sascha Hooker and Philip Clapham, and those from the 2003 Department of Fisheries and Oceans study are courtesy of Jack Lawson. Many thanks are extended to Shannon Gowans and anonymous reviewers for constructive comments on the manuscript.

References

Buckland, S.T., and Garthwaite, P.H. 1991. Quantifying precision of mark–recapture estimates using the bootstrap and related methods. Biometrics, **47**: 255–268.

Burnham, K.P., and Anderson, D.R. 2002. Model selection and multimodel inference: a practical information–theoretic approach. 2nd ed. Springer-Verlag, New York.

Burnham, K.P., Anderson, D.R., White, G.C., Brownie, C., and Pollock, K.H. 1987. Design and analysis of methods for fish survival experiments based on release–capture. American Fisheries Society, Bethesda, Md. Monograph 5.

Calambokidis, J., and Barlow, J. 2004. Abundance of blue and humpback whales in the Eastern North Pacific estimated by capture–recapture and line-transect methods. Mar. Mamm. Sci. **20**: 63–85.

Cormack, R.M. 1985. Examples of the use of GLIM to analyse capture–recapture studies. *In* Statistics in ornithology. *Edited by* B.J.T. Morgan and P.M. North. Springer-Verlag, Berlin. pp. 243–273.

COSEWIC. 2002. Update to the status of the northern bottlenose whale, *Hyperoodon ampullatus*, (Scotian Shelf population). Report of Committee on the Status of Endangered Wildlife in Canada, Ottawa, Ont.

Dalebout, M.L., Hooker, S.K., and Christensen, I. 2001. Genetic diversity and population structure among northern bottlenose whales, *Hyperoodon ampullatus*, in the western North Atlantic. Can. J. Zool. **79**: 478–484.

Gowans, S., and Whitehead, H. 2001. Photographic identification of northern bottlenose whales (*Hyperoodon ampullatus*): sources of heterogeneity from natural marks. Mar. Mamm. Sci. **17**: 76–93.

Gowans, S., Whitehead, H., Arch, J.K., and Hooker, S.K. 2000*a*. Population size and residency patterns of northern bottlenose whales (*Hyperoodon ampullatus*) using the Gully, Nova Scotia. J. Cetacean Res. Manag. **2**: 201–210.

Gowans, S., Dalebout, M.L., Hooker, S.K., and Whitehead, H. 2000*b*. Reliability of photographic and molecular techniques for sexing northern bottlenose whales (*Hyperoodon ampullatus*). Can. J. Zool. **78**: 1224–1229.

Gowans, S., Whitehead, H., and Hooker, S.K. 2001. Social organization in northern bottlenose whales (*Hyperoodon ampullatus*): not driven by deep water foraging? Anim. Behav. **62**: 369–377.

Hammond, P.S. 1986. Estimating the size of naturally marked whale populations using capture–recapture techniques. Rep. Int. Whaling Comm. Spec. Issue, **8**: 253–282.

Hammond, P.S. 1990. Heterogeneity in the Gulf of Maine? Estimating humpback whale population size when capture probabilities are not equal. Rep. Int. Whaling Comm. Spec. Issue, **12**: 135–139.

Pledger, S., Pollock, K.H., and Norris, J.L. 2003. Open capture–recapture models with heterogeneity: I. Cormack–Jolly–Seber model. Biometrics, **59**: 786–794.

Pollock, K.H. 1982. A capture–recapture design robust to unequal probability of capture. J. Wildl. Manag. **46**: 752–757.

Seber, G.A.F. 1982. The estimation of animal abundance and related parameters. 2nd ed. Griffin, London.

Seber, G.A.F. 1992. A review of estimating animal abundance II. Int. Stat. Rev. **60**: 129–166.

White, G.C., and Burnham, K.P. 1999. Program MARK: survival estimation from populations of marked animals. Bird Study, **46**(Suppl.): 120–138.

Whitehead, H. 1990. Mark–recapture estimates with emigration and re-immigration. Biometrics, **46**: 473–479.

Whitehead, H., Gowans, S., Faucher, A., and McCarrey, S. 1997. Population of northern bottlenose whales in the Gully, Nova Scotia. Mar. Mamm. Sci. **13**: 173–185.

Wimmer, T., and Whitehead, H. 2004. Movements and distribution of northern bottlenose whales, *Hyperoodon ampullatus*, on the Scotian Slope and in adjacent waters. Can. J. Zool. **82**: 1782–1794.

Zeh, J., Poole, D., Miller, G., Koski, W., Baraff, L., and Rugh, D. 2002. Survival of bowhead whales, *Balaena mysticetus*, estimated from 1981–1998 photoidentification data. Biometrics, **58**: 832–840.

Appendix A. Population estimation models.

There are J samples at times $t_1,\ldots,t_J$, and on the jth sample n_j individuals are identified. Let N_j be the population size at the jth sample. We assume that the population consists of two classes, A and B, with population sizes αN_j and $(1 - \alpha)N_j$ during the jth sample. The classes have different identification rates and mortalities.

If the population is growing at a rate r per year, and N_{M} is the population size at the midpoint of the sampling, $0.5(t_J + t_1)$, then

$$\text{(A1)}\quad N_j = N_{\mathrm{M}} e^{r[t_j - 0.5(t_J + t_1)]}$$

The estimated identification rate for the whole population in the jth sample is

$$\text{(A2)}\quad p_j = n_j / N_j$$

To represent differential identification rates for members of the two classes, $p_j(\mathrm{A})$ and $p_j(\mathrm{B})$, we use a logit link function (as in Pledger et al. 2003):

$$\text{(A3)}\quad \log\left[\frac{p_j(\mathrm{A})}{1 - p_j(\mathrm{A})}\right] = \mu_j$$

$$\text{(A4)}\quad \log\left[\frac{p_j(\mathrm{B})}{1 - p_j(\mathrm{B})}\right] = \mu_j + \nu$$

Equations A3 and A4 can be rewritten as

$$\text{(A5)}\quad p_j(\mathrm{A}) = \frac{1}{1+\lambda_j}$$

$$\text{(A6)}\quad p_j(\mathrm{B}) = \frac{1}{1+\beta\lambda_j}$$

where $\beta = e^{-\nu}$ and $\lambda_j = e^{-\mu_j}$. If there are very few identifications, then λ_j is large, so $p_j(\mathrm{A})/p_j(\mathrm{B}) \approx \beta$. Thus, β represents the ratio of identifiabilities of the members of the two classes. Now

$$\text{(A7)}\quad p_j = \alpha p_j(\mathrm{A}) + (1-\alpha)p_j(\mathrm{B})$$

Substituting expressions A2, A5, and A6 in eq. A7 gives

$$\text{(A8)}\quad n_j/N_j = \frac{\alpha}{1+\lambda_j} + \frac{1-\alpha}{1+\beta\lambda_j}$$

Solving for λ_j in eq. A8 (using the quadratic formula) gives

$$\text{(A9)}\quad \lambda_j = \frac{-b+\sqrt{b^2+4\beta(n_j/N_j)(1-n_j/N_j)}}{2\beta(n_j/N_j)}$$

where

$$\text{(A10)}\quad b = (n_j/N_j)(1+\beta) - \beta\alpha + \alpha - 1$$

Thus $p_j(\mathrm{A})$ and $p_j(\mathrm{B})$ can be expressed as functions of α, β, and n_j/N_j. These can be used to calculate the probability that a member of class A identified in sample j is next identified in sample k:

$$\text{(A11)}\quad q_{jk}(\mathrm{A}) = e^{-\delta_\mathrm{A}(t_k-t_j)}p_k(\mathrm{A})\prod_{x:j<x<k}[1-p_x(\mathrm{A})]$$

where δA is the mortality rate of class A. The probability that a member of class A identified in sample j and not identified again is

$$\text{(A12)}\quad s_j(\mathrm{A}) = \prod_{k:j<k\le J}\left\{e^{-\delta_\mathrm{A}(t_{k-1}-t_j)}[1-e^{-\delta_\mathrm{A}(t_k-t_{k-1})}]\prod_{x:j<x<k}[1-p_x(\mathrm{A})]\right\} + e^{-\delta_\mathrm{A}(t_J-t_j)}\prod_{k:j<k\le J}[1-p_k(\mathrm{A})]$$

We can write similar expressions for class B by substituting B for A in eqs. A11 and A12.

Then the log-likelihood that an individual i has an identification history $t_{i,1}, t_{i,2},\ldots,t_{i,I}$, conditioning on its first identification ($t_{i,1}$) is

$$\text{(A13)}\quad L(t_{i,1},t_{i,2},\ldots,t_{i,I}) = \sum_{j=1,\ldots,I-1}\log[\alpha q_{t_{i,j}t_{i,j+1}}(\mathrm{A}) + (1-\alpha)q_{t_{i,j}t_{i,j+1}}(\mathrm{B})] + \log[\alpha s_{t_{i,I}}(\mathrm{A}) + (1-\alpha)s_{t_{i,I}}(\mathrm{B})]$$

The log-likelihood of the data set is then

$$\text{(A14)}\quad L = \sum_i\left\{\sum_{j=1,\ldots,I-1}\log[\alpha q_{t_{i,j}t_{i,j+1}}(\mathrm{A}) + (1-\alpha)q_{t_{i,j}t_{i,j+1}}(\mathrm{B})] + \log[\alpha s_{t_{i,I}}(\mathrm{A}) + (1-\alpha)s_{t_{i,I}}(\mathrm{B})]\right\}$$

The population parameters N_M, r, α, β, δ_A, and δ_B are then chosen to maximize expression A14 using the Nelder–Mead Simplex method.

Various constraints on the parameters permit simplifications of this full model: $r = 0$: stable population size, no trend; $\alpha = 1$: one class of individual, no heterogeneity; $\beta = 1$: no heterogeneity in identification rates; $\delta_\mathrm{A} = \delta_\mathrm{B}$: no heterogeneity in mortality; $\delta_\mathrm{A} = \delta_\mathrm{B} = 0$, $\alpha = 1$: no mortality or heterogeneity, closed Schnabel model.

Using Bayes' theorem, it is possible to provide a posterior allocation of the individuals into the two classes, A and B (Pledger et al. 2003). The probability that individual i with catch history $t_{i,1}, t_{i,2},\ldots, t_{i,I}$ was a member of class A is

$$\text{(A15)}\quad \Pr(\mathrm{A}\,|\,t_{i,1},t_{i,2},\ldots,t_{i,I}) = \frac{\alpha s_{t_{i,I}}(\mathrm{A})\prod_{j=1,\ldots,I-1}q_{t_{i,j}t_{i,j+1}}(\mathrm{A})}{\alpha s_{t_{i,I}}(\mathrm{A})\prod_{j=1,\ldots,I-1}q_{t_{i,j}t_{i,j+1}}(\mathrm{A}) + (1-\alpha)s_{t_{i,I}}(\mathrm{B})\prod_{j=1,\ldots,I-1}q_{t_{i,j}t_{i,j+1}}(\mathrm{B})}$$

The probability that it was a member of class B is then $1 - \Pr(\mathrm{A}\,|\,t_{i,1}, t_{i,2},\ldots,t_{i,I})$.

Appendix B. Simulated data sets for parametric bootstraps.

To construct simulated data sets using the "heterogeneity in identifiability" model, we took the parameter estimates of N_M, δ, β, and α for the left-side identifications using this model from Table 4. N_M individuals formed the population at the sampling period 1. Individuals had a probability $e^{-\delta(t_{j+1}-t_j)}$ of surviving from the jth to the (j + 1)th sampling period. Those dying were replaced by new individuals. All individuals in the population had a probability α of belonging to class A and $1 - \alpha$ of belonging to class B. The probability that an individual in the population in the jth sample was identified during that sample was given by eqs. A5–A10. This produced a simulated set of identification histories that could be input into the parameter estimation routines, as described in Appendix A.

To construct simulated data sets using the "heterogeneity in mortality & identifiability" model, we gave the members of the two classes different probabilities of survival between sampling periods (calculated using estimates of δ_A and δ_B). Dead individuals were replaced by individuals from their own class, thus maintaining the relative contributions of each class to the population.

Simulated data sets with continuous rather than bimodal distributions of identifiability were constructed by giving individual i an identifiability determined by $y_i = 1 - (1 - \gamma)\sqrt{u}$, where u is a uniform 0–1 random variable and γ (= 0.01) was chosen so that the variance in identification rates was similar to that in the bimodal case with $\alpha = 0.76$ and $\beta = 0.11$ (as estimated using the left-side identifications and the "heterogeneity in identifiability" model; Table 4). The distributions of identifiability in this case and that of the bimodal model are compared (Fig. B1). With continuous identifiability, eqs. A5 – A10 could not be used to estimate identification rates for each sampling period. However, generalizing eq. A8, the overall identification rate for sampling period j is given by

$$\text{(B1)} \quad n_j / N_j = \left(\sum_i \frac{1}{1 + y_i \lambda_j} \right) / N_j$$

For each sampling period j, λ_j was calculated by solving eq. B1 (numerically using the fzero function of MATLAB, The MathWorks Inc., Natick, Massachusetts), and so the probability of identification for individual i in sampling period j was given by $1/(1 + y_i\lambda_j)$. This allowed the construction of simulated data sets, as in the cases described above.

Fig. B1. Distributions of identifiability with the bimodal model (Class A and Class B; solid lines) and with continuous identifiability (dashed lines).