Minke whale abundance estimation from the NASS 1987 and 2001 aerial cue–counting surveys taking appropriate account of distance estimation errors

David L. Borchers[1], Daniel G. Pike[2], Thorvaldur Gunnlaugsson[3] and Gísli A. Víkingsson[3]

[1] *RUWPA, The Observatory, University of St Andrews, Fife, KY16 9LZ, United Kingdom*
[2] *1210 Ski Club Road, North Bay, Ontario, P1B 8E5, Canada*
[3] *Marine Research Institute, P.O. Box 1390, IS-121, Reykjavík, Iceland*

ABSTRACT

We estimate the abundance of minke whales (*Balaenoptera acutorostrata*) from the Icelandic coastal shelf aerial surveys carried out as part of the 1987 and 2001 North Atlantic Sightings Surveys (NASS). In the case of the 1987 survey, the probability of detecting animals at distance zero ($g(0)$) is very close to 1 but there is substantial random measurement error in estimating distances. To estimate abundance from these data, we use methods which assume $g(0)=1$ but which include a distance measurement error model. In the case of the 2001 survey, measurement errors were sufficiently small to be negligible, and we use double platform methods which estimate $g(0)$ and assume no measurement error to estimate abundance. From the 1987 survey, we estimate abundance to be 24,532 animals, with 95% CI (13,399; 44,916). From the 2001 NASS survey data, minke whale abundance is estimated to be 43,633 animals, with 95% CI (30,148; 63,149).

INTRODUCTION

The first aerial survey using cue–counting techniques in Icelandic waters was conducted in 1987 (Donovan and Gunnlaugsson 1989), and surveys using almost identical designs were conducted in 1995 and in 2001 as part of the North Atlantic Sightings Survey (NASS) programme. The history and development of these surveys is fully described by Pike *et al.* (2009b). The primary target species of the surveys has been the minke whale (*Balaenoptera acutorostrata*) and it is for this species that cue–counting has proven most effective. In this methodology, described by Hiby and Hammond (1989), indicators of whale presence ("cues"), such as blows or surfacings in the case of minke whales, are counted, rather than animals or groups of animals. The method has an important advantage over line transect techniques in that it involves a weaker assumption about detection at distance zero. Rather than assuming that all animals along the trackline are seen, it is assumed that all cues at radial distance zero are seen. The resulting estimate of cue density per unit time is converted to whale density using an estimate of cue rate. Cue–counting estimates are thus much less subject to availability bias (bias due to animals being unavailable for detection). Double platform data may however be needed to correct for perception bias (bias due to missing available animals). An estimate of the cue rate of the target species is required, and this can be obtained by tagging or observational experiments.

Pike *et al.* (2009b) provide estimates of relative abundance of minke whales from these aerial surveys based on standard line transect analyses. These estimates are however negatively biased because they do not correct for

availability and perception biases. Estimates of absolute abundance from the 1987 survey have been produced by Hiby *et al.* (1989) and Borchers *et al.* (MS 1997); the latter authors also provided an estimate for the 1995 survey. The NAMMCO Scientific Committee later concluded that deriving a reliable estimate from the 1995 survey was problematic because measurement error and perception bias could not be assessed (NAMMCO 2002). Therefore data from that survey will not be analyzed here.

The 2 previous analyses produced very different abundance estimates for the 1987 survey. Borchers *et al.* (MS 1997) used standard cue–counting analysis methods while Hiby *et al.* (1989) used a method that incorporated measurement error. The latter survey focused attention on the potentially large bias that measurement error can induce in the analysis of cue–counting surveys. The method developed by Hiby *et al.* (1989) for estimating detection probability and abundance from cue–counting surveys in the presence of measurement error uses distance data grouped into intervals. In this paper we use maximum likelihood estimators (MLEs) based on ungrouped distance data. The MLE of detection probability using a variety of detection function forms were simulation tested by Borchers *et al.* (submitted), who found it to be approximately unbiased over all scenarios investigated, while conventional distance sampling estimators were found to be substantially biased by measurement errors when the coefficient of variation (cv) of measurement error is not small (greater than about 10%).

We use a maximum-likelihood method which incorporates measurement error to estimate minke whale abundance from NASS 1987 data, which have large distance measurement errors. In the case of the NASS 2001 data, measurement errors are sufficiently small that cue–counting methods which do not incorporate these errors can be used without fear of substantial bias. We compare our results for the 1987 survey to earlier analyses, and suggest reasons for discrepancies. We also examine the trend in absolute abundance of minke whale in Icelandic waters between 1987 and 2001, and suggest reasons for the observed changes.

METHODS

1987 survey
Details of the survey methods and design are given in Donovan and Gunnlaugsson (1989) and Pike *et al.* (2009b), and the survey area, stratification and planned tracklines are shown in Fig. 1.
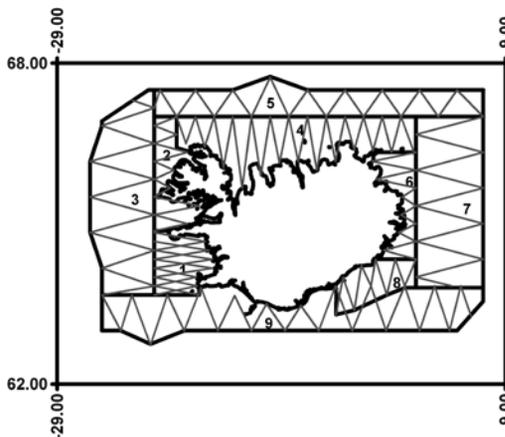
Fig. 1. *Stratification and planned survey effort for the 1987 and 2001 surveys. The numbers refer to the strata.*

Salient features of the data for the current analysis include the following.

(1) Double platform survey was conducted only on part of 1 day, generating only 7 duplicates. There were a further 8 duplicates with independent estimates of distance but with dependent detection. Hiby *et al.* (1989) estimated $p(0)$ for platform 1 to be 0.97 and that for platform 2 to be 0.91. This implies that the probability of either platform detecting an animal on the trackline is 0.997.

(2) Some periods of survey were conducted with compromised effort on 1 side of the aircraft. Effort was compromised by equipment problems which resulted in observers being unable to search for all the time during which they were nominally searching. In these cases on data from the uncompromised side have been used (with a corresponding halving of the region searched).

(3) Distances to duplicate detections indicate that there were sometimes large errors made in estimating distances, in both directions (underestimation and overestimation).

Because $p(0)$ is so close to 1 for the combined platform, the data are analysed here as data from a single combined platform assuming $p(0)=1$.

2001 survey

The survey methods and design are described in detail in Pike *et al* (2009b). The primary observers were located in the left and right rear seats of the aircraft, searching through bubble windows. For brevity we denote them P1 and P2. The cruise leader (denoted S1), sitting in the right front seat, is treated as a secondary observer.

Salient features of the data are as follows:

(1) Primary and secondary observers on the right side of the plane were separated by a curtain and aural independence was maintained as far as possible. Independence between the primary observer on the left side of the plane and the pilot was compromised because the pilot was not visually isolated from the observer behind him. Distance estimates by the pilot are also not reliable due to timing problems.

(2) Double platform survey was maintained throughout the survey, but because of lack of independence on the left side of the plane, only double platform data from the right side of the plane are treated as such. P1 and P2 switched sides at least daily.

(3) Distances to duplicate detections indicated small errors in estimating distances.

Modelling and statistical methods

Conventional cue–counting methods are based on distances to detections being observed without error. In many applications errors are small enough that this is a reasonable approach. In this case, the cue–counting likelihood function involves only a model for detection probability as a function of distance (and possibly other variables), which we denote $p(\ )$, and a model for the distribution of distances of animals in the vicinity of the observer, which we denote $\pi(\ )$. When measurement errors are incorporated, an appropriate likelihood involves a measurement error model as well, which we denote $m(\ )$. Below we develop measurement error models and then incorporate them in likelihood functions

which have components involving $p(\ )$, $\pi(\ )$ and $m(\ )$. Estimation using these methods was implemented in the statistical programming language R (The R Foundation for Statistical Computing 2008) and maximisation of the likelihood functions was done with the R function optim( ) using a Nelder-Mead optimisation algorithm. This function provides diagnostics relating to convergence, which in all cases here indicated successful convergence. Convergence was also checked by using a number of different starting values for parameters being estimated.

Measurement error model

Borchers (MS 2003) investigated models with homoscedastic normal errors and heteroscedastic normal errors with standard error proportional to true distance. They also noted that a gamma error model provides a flexible alternative to a normal model. In addition, for aerial surveys homoscedastic error models seem implausible at small distances, because it is unlikely that observers will estimate a detection to be on the opposite side of the observer to the sighting, but a normal error model does not preclude this. In this paper, we model distance estimation errors primarily using the gamma distribution and we assume unbiased distance estimation (Fig. 3 below suggests this is not unreasonable).

We call the true radial distance from the observer to a detected object R, and the estimated radial distance to it r. If there is no measurement error, then $r=R$. We model measurement error via the probability density function (pdf) of r, given R and assume either a multiplicative normal or a gamma form for the errors. We denote the pdf of r given R $m(r|R)$, which has parameter vector ϕ. In the case of a normal error model $\phi = \sigma^2$ and $r=R\epsilon$, where ϵ is a normal random variable with mean 1 and variance σ^2. In the case of a gamma error model, (1)

$$m(r \mid R) = \left[\left(\frac{R}{\alpha} \right)^{\alpha} \Gamma(\alpha) \right]^{-1} r^{\alpha-1} \exp \left\{ -\frac{\alpha r}{R} \right\}$$

$\phi = \alpha$, $E[r]=R$ and the coefficient of variation (cv) of r is $1/\sqrt{\alpha}$.

Single and combined platform detection function likelihood

Let $\pi(R)$ be the probability density function (pdf) of the distance R of an animal (whether detected or not) from the observer(s). The probability of detecting an animal depends on R and a vector of other variables which we denote z. We use the $p()$ to denote detection probability functions as follows: $p(R,z)$ is the probability of detecting an animal at radial distance R, with explanatory variables z, and $p(\underline{z}) = \int_0^\infty p(R,\underline{z})\pi(R)dR$ is the probability of detecting an animal with explanatory variables $\underline{z}$. We use Θ to denote the parameters of $p(R,z)$.

The pdf of the recorded distance r obtained when explanatory variables z apply, is

(2)
$$f(r|\underline{z}) = \int_0^\infty m(r|R)f(R|\underline{z})dR$$
$$= \frac{\int_0^\infty m(r|R)p(R,\underline{z})\pi(R)dR}{p(\underline{z})}$$

It follows that the likelihood for the parameter vectors Θ and ϕ, given the n distances $r_1,...,r_n$ that were recorded by is

(3)
$$L(\theta,\underline{\phi}|r_1,...,r_n) = \prod_{i=1}^n f(r_i|\underline{z})$$

When the parameter ϕ of the pdf $m(r|R)$ is known, Θ can be estimated by maximizing this likelihood with respect to Θ. Without additional data on the measurement error process, Θ cannot be estimated from single observer data.

Estimation of measurement error model parameter(s)

If there are supplementary data available for which true distances are known, the parameter vector ϕ of the measurement error model $m(r|R)$ can be estimated from these data. If true distances are unknown but independent pairs of estimated distances to the same animals are available, it may still be possible to estimate ϕ from these data. For the 1987 survey, for example, there are 15 pairs of estimated distances to the same animals (a pair comprising independent estimates from each of the 2 platforms).

Assuming that the 2 platforms are subject to the same measurement error model $m(r|R)$ and that measurements are unbiased, ϕ can be estimated by maximum likelihood. When there are K such radial distance estimates available, the likelihood function for is as follows:

(4)
$$L(\underline{\phi},R_1,...,R_K|(r_{11},r_{21}),...,(r_{1K},r_{2K})) = \prod_{k=1}^K m(r_{1k}|\underline{R_k})m(r_{2k}|\underline{R_k})$$

Maximum likelihood estimators (MLEs) are not necessarily unbiased for small samples and simulation testing of the MLE of ϕ with $K=15$ and measurement error cv between 0.1 and 0.5 indicated that the MLE of the function cv(ϕ) can be substantially negatively biased (greater than 40% in some cases). (For the normal error model cv(ϕ)$= \sigma$ and for the gamma error model cv(ϕ)$= 1/\sqrt{\alpha}$) We therefore implemented an iterative bias corrected estimation procedure as follows;

1. Initialise the bias correction factor b
2. Obtain the MLEs $(\hat{\underline{\phi}},\hat{R}_1,...,\hat{R}_K)$ and hence cv(ϕ) by maximising Equation (4).
3. Simulate D sets of radial distance pairs $(r_{11},r_{21}),...,(r_{1K},r_{2K})$ from $m(r|R)$ using parameters $((b \times \hat{\underline{\phi}}),\hat{R}_1,...,\hat{R}_K)$.
4. Obtain MLEs $(\hat{\underline{\phi}},\hat{R}_1,...,\hat{R}_K)_d$ and cv(ϕ)$_d$ (d=1,...,D) from the D simulated data sets and estimate a new bias correction factor $b=$ cv(ϕ)/ mean(cv(ϕ)$_d$)
5. If the new b from step 4 above is within δ% of the previous b go to step 6, else go to step 2 above.
6. Estimate ϕ as $(b \times \hat{\underline{\phi}})$.

In our application of this bias correction algorithm we used D=1000 and δ=1 (*i.e.* convergence was assumed when b changed by less than 1% between iterations).

Double platform detection function likelihood
We do not deal with likelihoods involving both mark recapture cue–counting data and measurement error in this paper. For the 1987 data we use a method which incorporates measurement error but no mark recapture component. In the case of the 2001 data we use a mark recapture distance sampling (MRDS) method similar to the method of Borchers *et al.* (2006), without measurement error. This latter method involves estimation of the probability

Table 1. 1987 survey. The estimated abundance and related parameters from the strata of Donovan and Gunnlaugsson (1989) using an MCDS model with Beaufort Sea State. Measurement error cv was estimated to be 32%, with standard error 4%. $\hat{h}(0)$ is the estimated slope of the pdf of radial distances at the origin. (Although N was estimated in only 6 strata, detections from all 7 strata were used to estimate $h(0)$.) Areas are in nm²; time (T) is in hours.

Stratum	Area	T	n	n/T	$\hat{h}(0)$	$\hat{N}$	%cv	95%CI
1	3,734	7.29	57	7.81	24.2	5,537	30	(3,086; 9,931)
2	3,928	4.47	12	2.68	24.5	2,124	61	(706; 6,388)
4	12,357	14.06	24	1.70	13.3	1,732	32	(932; 3,220)
5	6,944	3.15	3	0.95	21.6	845	55	(309; 2,311)
8	1,659	2.81	34	12.09	18.2	2,295	37	(1,133; 4,650)
9	16,617	7.02	13	1.99	19.7	3,933	62	(1,289; 12,019)
Total	**45,239**	**39.01**	**144**	**3.64**	**21.9**	**16,468**	**36**	**(8,287; 32,724)**

Table 2. 1987 survey. The estimated abundance and related parameters from the original design strata using an MCDS model with Beaufort Sea State. Measurement error cv was estimated to be 32%, with standard error 4%. $\hat{h}(0)$ is the estimated slope of the pdf of radial distances at the origin. Areas are in nm²; time (T) is in hours.

Stratum	Area	T	n	n/T	$\hat{h}(0)$	$\hat{N}$	%cv	95%CI
1	4,418	7.29	57	7.81	24.2	6,572	29	(3,771; 11,452)
2	3,988	4.47	12	2.68	24.5	2,163	64	(690; 6,784)
3	14,066	0.96	1	1.04	37.1	3,587	110	(623; 20,653)
4	12,392	14.06	24	1.71	13.3	1,743	32	(951; 3,194)
5	9,471	3.15	3	0.95	21.6	1,156	62	(380; 3,515)
8	3,728	2.81	34	12.09	18.2	5,174	36	(2,613; 10,245)
9	17,408	7.02	14	1.99	19.7	4,137	61	(1,387; 12,337)
Total	**65,471**	**39.98**	**145**	**3.65**	**21.9**	**24,532**	**32**	**(13,399; 44,916)**

of detecting a cue at radial distance zero from double observer data, and separate estimation of the shape of the radial distance detection function using multiple covariate distance sampling (MCDS) methods, as implemented in the programme Distance (Thomas *et al.* 2005).

Detection function models
In conventional distance sampling (CDS), detection at zero distance is certain ($p(0)=1$) and detection probability is modelled as a function of distances alone. The "pooling robustness" property of CDS estimators means that neglecting variables other than distance which affect detection probability does not introduce bias. MCDS models retain the $p(0)=1$ assumption but include covariates z other than distance, which may affect detection probability. When detection at zero distance might be less than certain, MRDS detection probability models are appropriate.

These allow uncertain detection at distance zero and include covariates other than distance; mark recapture estimators are not "pooling robust" so including all covariates that may affect detection probability is important for unbiased estimation.

Abundance estimation
We consider estimation for the case in which the survey region is divided into S geographical strata (numbered $s=1,\ldots,S$). Observers search out to radial distance W but they may search a variable fraction of the circle about them as they proceed along the transect lines. Suppose that in stratum s they search the same fraction of the circle, Υ_k for a time T_{sk} ($k=1,\ldots,K$), so that the area about the observer that is searched on the kth of these sections in stratum s is $\Upsilon_k \pi W^2$.

We estimate the density of cues per unit time in stratum s by

(5)
$$\hat{C}_s = \sum_{k=1}^{K_s} \sum_{i=1}^{n_{sk}} \frac{1}{\hat{p}(\underline{z}_i) T_{sk} a_{sk}}$$

where a_{sk} is the surface area searched on the kth section in stratum s,

(6)
$$\hat{p}(\underline{z}_i) = \int_0^W [\hat{p}_1(R, \underline{z}_i) + \hat{p}_2(R, \underline{z}_i) - \hat{p}_1(R, \underline{z}_i)\hat{p}_2(R, \underline{z}_i)]\pi(R)dR$$

if 2 platforms were operating independently ("hats" indicate estimates and subscripts 1 and 2 indicate platforms 1 and 2 respectively), and

(7)
$$\hat{p}(\underline{z}_i) = \int_0^W \hat{p}_\bullet(R, \underline{z}_i)\pi(R)dR$$

if the 2 platforms are considered a single team (subscript "•" indicates platforms 1 and 2 combined). Here W is a truncation distance sufficiently large to include all observed objects used in the analysis, n_{sk} is the number of different cues detected in on the kth segment of transects in stratum s and $\pi(R)=2R|W^2$. The resulting estimator of animal abundance is

(8)
$$\hat{N} = \sum_{s=1}^{S} \hat{N}_s = \sum_{s=1}^{S} \frac{\hat{C}_s}{\hat{\lambda}} A_s$$

where A_s is the surface area of stratum s and $\hat{\lambda}$ is an estimate of the mean cue rate the mean number of cues produced per animal per time unit. A cue rate of 53 cues per hour per animal (with zero variance) was assumed for minke whales (Gunnlaugsson 1989). In general, a different mean cue rate might be estimated for each stratum, but here we consider the case in which a single cue rate is estimated for all strata.

When animals occur in groups, the above estimator of cue density is modified as follows

(9)
$$\hat{C}_s = \sum_{k=1}^{K_s} \sum_{i=1}^{n_{sk}} \frac{z_i}{p(\underline{z}_i) T_{sk} a_{sk}}$$

where z_i is the size of the ith group. Mean school size is estimated as the ratio of the abundance estimate obtained using Equation (9), and that obtained using Equation (5).

Estimation of h(0)
Detection probability estimates are conventionally presented in terms of $h(0)$, the slope of the pdf of observed distances, evaluated at distance $R=0$. When detection probability depends on the vector of variables, z, we write this as $h(0|z)$, which is equal to $2/[p(z)W^2]$. For comparability with conventional distance sampling estimates of $h(0)$ (which do not depend on z), we present below estimates of the mean $h(0|z)$ in each stratum. This is a weighted average of the $h(0|z_i)$s of each detected cue in the stratum:

(10)
$$\hat{h}_s(0) = \sum_{k=1}^{K_s} \sum_{i=1}^{n_{sk}} \frac{h(0 \mid \underline{z}_i)}{p(\underline{z}_i) T_{sk} a_{sk} C_s}$$

The overall $h(0)$ across strata is estimated similarly:

(11)
$$\hat{h}(0) = \sum_{s=1}^{S} \hat{h}_s(0) \frac{\hat{N}_s}{\hat{N}}$$

Variance and confidence interval estimation
Variance is estimated by a 2 step bootstrap procedure as follows:

1. Draw a nonparametric bootstrap sample using the transects as the re-sampling unit and re-sampling separately within each stratum.
2. Within each nonparametric bootstrap sample, draw a parameteric bootstrap sample comprising K pairs of independent radial distance estimates from the error model with parameters $((b \times \hat{\phi}), \hat{R}_1,..., \hat{R}_K)$.
3. Obtain estimates $(\hat{\phi}, \hat{R}_1,..., \hat{R}_K)$ from the sample in 2. above by maximising Equation (4) above.
4. Estimate abundance and related parameters by maximising Equation (3), given the estimates $(\hat{\phi}, \hat{R}_1,..., \hat{R}_K)$ from 3. above, after correcting the bias in $\hat{\phi}$ by multiplying it by b. (The bias correction factor b was estimated as described above from the original sample and was not re estimated within the bootstrap procedure.)
5. If fewer than 999 bootstrap samples have been drawn, go to 1. above, else go to 6 below.
6. Calculate 95% log based confidence intervals, as described in Buckland *et al.* (2001: p77).

Estimator properties
Borchers *et al.* (MS 1997) and Borchers *et al.* (submitted) investigated the properties of the $h(0)$ estimator above for single platform surveys of line transect and point transect surveys, without covariates, for a variety of detection

function shapes and for gamma measurement error cv's of 10%, 30% and 50%. The expected biases of the CDS estimators for the 10%, 30% and 50% cv scenarios were found to be 2%, 21% and 78%, respectively. Borchers *et al.* (submitted) also show that the *h(0)* estimator above is asymptotically unbiased.

RESULTS

Realized effort and distribution of sightings
Realized survey effort by block and the distribution of minke whale groups is shown in Donovan and Gunnlaugsson (1989) for the 1987 survey and by Pike *et al.* (2009b) for both 1987 and 2001, and is also shown in Fig. 2.

The distribution of minke whales was consistent between the 2 surveys. The areas of highest density were Block 1 (Faxaflói Bay, SW Iceland), Block 8 (SE Iceland) and Block 4 (N Iceland). Relatively few minke whales were seen in the offshore blocks.

1987 survey
Hiby *et al.* (1989) estimated measurement error coefficient of variation (cv) to be 35% using a normal error model. We estimated measurement error model parameters using the bias-corrected MLE method described above. This was done using normal and gamma measurement error models. In both cases the cv was estimated to be 32%. The gamma error models produced a substantially lower AIC than the normal model (-61 vs -53) and is preferred on this basis. The bias correction factor *b* was estimated to be 1.42. The cv of the gamma based measurement error cv estimate was estimated to be 13% using the inverse of the estimated information matrix obtained in maximising the likelihood.

While the pairs of estimated distances to duplicate detections can be used to estimate stochastic errors in distance estimation under the assumption of zero bias (and the assumption that distance estimation errors by all observers are governed by the same stochastic process), they are inadequate to estimate distance estimation bias. At the suggestion of a referee, we examined the estimated distance estimation errors by individual observer to investigate possible differences in distance estima-

tion bias between observers. The results are shown in Fig. 3, from which we conclude that there is inadequate evidence of differential bias between observers for this to be a concern.

Unless otherwise stated, the estimates below were obtained using a gamma error measurement error model.
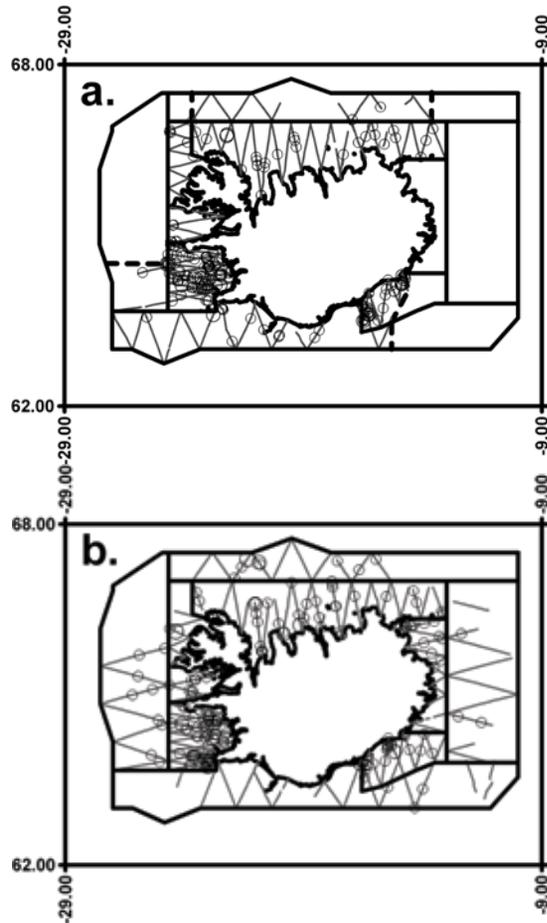
Fig. 2. Realized survey effort and sightings of minke whales in a) 1987, and b) 2001. Symbol size increases with group size in the range of 1 to 3. The strata of Donovan and Gunnlaugsson (1989) are shown by the dashed lines.

For comparison with the estimate of Hiby *et al.* (1989), who used a hazard rate model, we estimated abundance using Equation (2), with an MCDS hazard rate model, with scale parameter a function of Beaufort sea state, and with a 35% multiplicative normal measurement error cv (assumed, not estimated). We did not use data from compromised effort (Hiby *et*

al. (1989) might have), and used the strata of Donovan and Gunnlaugsson (1989) (Fig. 2a). This gives an estimated abundance of 13,246 whales (cv=39%) – compared to the estimate of Hiby *et al.* (1989) of 8,645 whales (cv=20%).

When we fit the hazard rate MCDS model with zero measurement error, the estimate increases by about 50% to 20,048 whales (cv=22%).

Best estimate
To obtain the best estimate from the current dataset, we used a gamma measurement error model and MCDS detection function models with scale parameters depending on the following explanatory variables (in addition to radial distance) were considered: Beaufort sea state, cloud cover, group size, glare, sightability index, position of observer (front or back) and heading relative to the aircraft. Half normal and hazard rate MCDS models were fitted. Selecting variables based on AIC resulted in the selection of a half normal detection function model with Beaufort sea state and radial distance as explanatory variables and a gamma error model.

The estimated detection functions and pdfs are shown in Figs 4 and 5. Note that one cannot evaluate the goodness of fit of the detec-
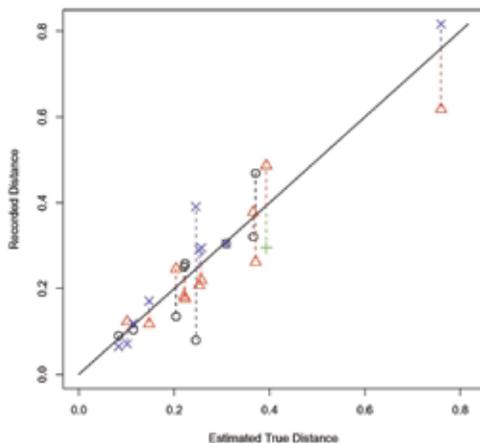
Fig. 3. *Observer estimates of distances to duplicate detections (vertical axis) plotted against the maximum likelihood estimate of distances to duplicates (horizontal axis) on the 1987 survey. Each observer is plotted with a different symbol and estimates to the same duplicate detection are linked by dashed lines. The solid diagonal line corresponds to the observer distance estimate being equal to the maximum likelihood estimate.*

tion function in the presence of measurement errors by their apparent fit to the histogram. This is because the distances in the histogram include measurement errors while the detection functions do not. The model with measurement error has a very much lower AIC than the model with no measurement error (-97 vs -58).

Abundance estimates and other parameters from this model, using the strata of Donovan and Gunnlaugsson (1989) are shown in Table 1. Estimates using the original strata are shown in Table 2. Mean school size is estimated, as described above, to be 1.10 (cv=8%). Total abundance using the strata of Donovan and Gunnlaugsson (1989) is estimated to be 16,468 whales (cv=36%). Note that all detections (including 1 in stratum 3) were used to estimate the detection function and mean school size, but encounter rate (n/T) data from stratum 3 were not used to estimate abundance with the strata of Donovan and Gunnlaugsson (1989). Total abundance using the original strata (including stratum 3) is estimated to be 24,532 whales (cv=32%).

2001 survey
Preliminary analysis of these data identified a potential problem with data from observer P2. This can be seen in Fig. 6: the histogram of duplicate proportions suggests that detection probability is low at the origin. However, sample size is small and the observer paired with P2 (observer S1) generated only 2 detections within the first 0.15 nm.

The 20 duplicates with reliable distance estimates were used to estimate the extent of distance measurement error. The cv of distance estimation error was estimated to be 11.6%. Measurement error this small has negligible effect on abundance estimates (Borchers *et al.* submitted). Therefore methods assuming no measurement error were used for estimation.

Estimation of p(0) and abundance
The methods of Borchers *et al.* (1998) were used to estimate *p*(0), the detection probability at distance zero, for all 3 observers operating on the right of the plane. Logistic detection function forms were assumed and explanatory variables were chosen using AIC from: Beaufort sea state, cloud cover, group size, glare, sight-
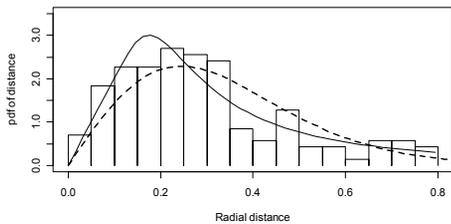
Fig. 4. *1987 Survey. Estimated probability density functions $\hat{f}(R)$. The histogram is the observed data (with measurement error). The solid curve is from the best MCDS model with no measurement error (hazard-rate form) including Beaufort sea state as a covariate and; the dashed curve is from the best MCDS model (half-normal form) with measurement error and including Beaufort sea state as a covariate.*
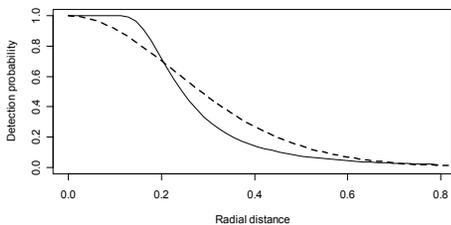
Fig. 5. *1987 Survey. Estimated hazard rate detection function models ($\hat{p}(R)$). The solid curve is from the best MCDS model with no measurement error (hazard-rate form) including Beaufort sea state as a covariate and; the dashed curve is from the best MCDS (half-normal form) with measurement error and including Beaufort sea state as a covariate.*

ability index and heading relative to the aircraft. All models include radial distance as models without radial distance as an explanatory variable were considered implausible. Models with radial distance and Beaufort were selected for estimation for all observers. Note that in the case of S1, detection probability increases with distance. However, S1 was in the front seat and had a somewhat obscured view of the trackline.

Figs 6 to 8 show the data used for estimation. Figs 9 to 11 show the estimated logistic functions for each of the 3 observers. Fig.12 shows the same for P2 and P1 treated as a single platform (with no observer effect in the model). Table 3 shows the estimates of detection probability at the origin (the intercept of the logistic detection functions) for each observer set.

In the case of observer P1, detection at distance zero is virtually certain ($\hat{p}_1(0) = 0.97$, cv=15%) and not significantly different from 1. CDS methods are therefore used to estimate abundance; this gives an estimate of 43,600 whales (cv=19%) using only data from the side on which P1 operated (see Table 4 and Figs 13 and 14). Table 5 summarises the cue–counting estimates of abundance when data from all observers is used; Figs 15 and 16 show the fitted logistic function and pdf.

Observer P2 certainly misses cues at zero distance, and although the point estimate for $p_2(0)$ is only 0.24, there is enormous uncertainty about this estimate (cv=86%). The cost in terms of variance of estimating $p_2(0)$ far outweighs the (possible) gain in terms of bias (estimated to be only 3% since $\hat{p}_1(0) = 0.97$) so that including estimation of $p_2(0)$ will result in a worse estimator (in terms of mean squared error of estimation) than neglecting it. We therefore do not include it.

Observer S1 generates about as many trials for P2 as for P1, and P2 and P1 operate for equal times on the survey, so the combined estimate of $p\bullet(0)$ for (P1+P2) provides a reasonable estimate of the mean probability of a primary observer (P2 or P1) detecting a cue at distance zero on the survey. Using the $p\bullet(0)$ estimate of 0.78 (cv=27%) to correct the estimate of 33,600 obtained using CDS methods with all the data (from the side with P2 and the side with P1), gives an estimate of 43,100 (cv=32%). This corrected point estimate agrees very well with the estimate obtained using data from the side with P1 alone (it differs by about 1%), but it has a higher cv, even though it is obtained using more data. This is because it involves estimation of $p\bullet(0)$ and cv[$p\bullet(0)$] is relatively high (27%). So somewhat counter-intuitively, discarding data from the side with observer P2 provides a more

Table 3. 2001 survey. The estimated probability of detection at the origin, $\hat{p}(0)$, for observers P2, P1, S1 and for P2 and P1 combined (P1+P2).

Observer	$\hat{p}(0)$	%cv	95% CI
P1	0.97	15%	(0.61; 1.00)
P2	0.24	86%	(0.002; 0.80)
P1+P2	0.78	27%	(0.25; 0.97)
S1	0.14	50%	(0.02; 0.30)
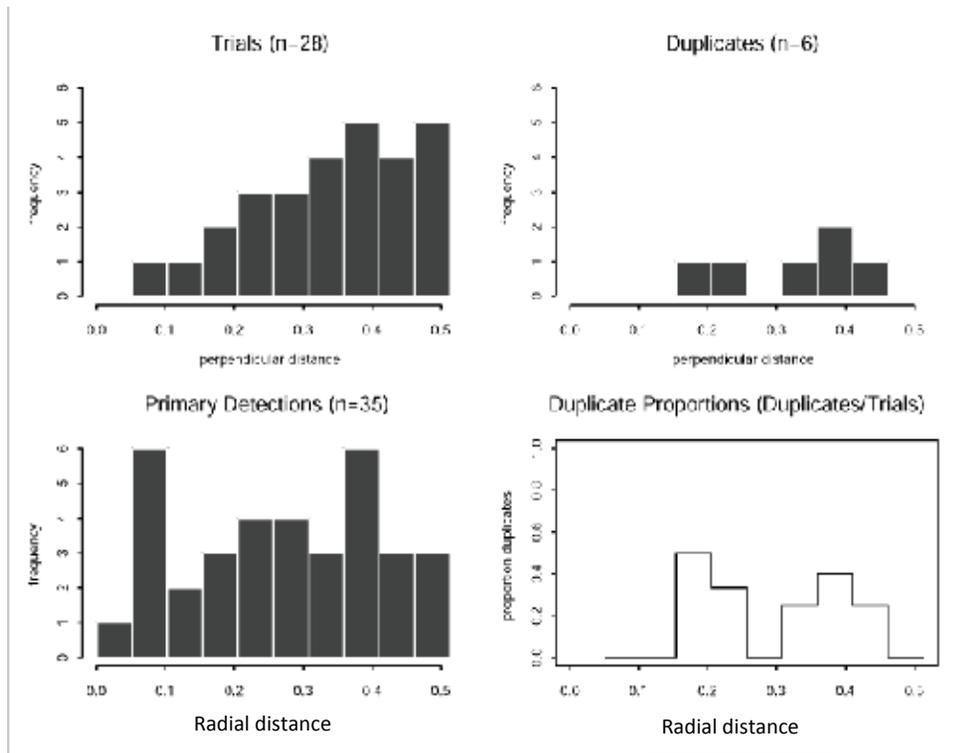
Fig. 6. *2001 Survey. Detections by observer S1 ("Trials": top left plot), observer P2 ("Primary Detections": bottom left plot), and duplicates (top right plot) for periods when S1 and P2 were paired on the right of the aircraft. The bottom right plot shows the proportions of S1 detections which were detected by P2.*
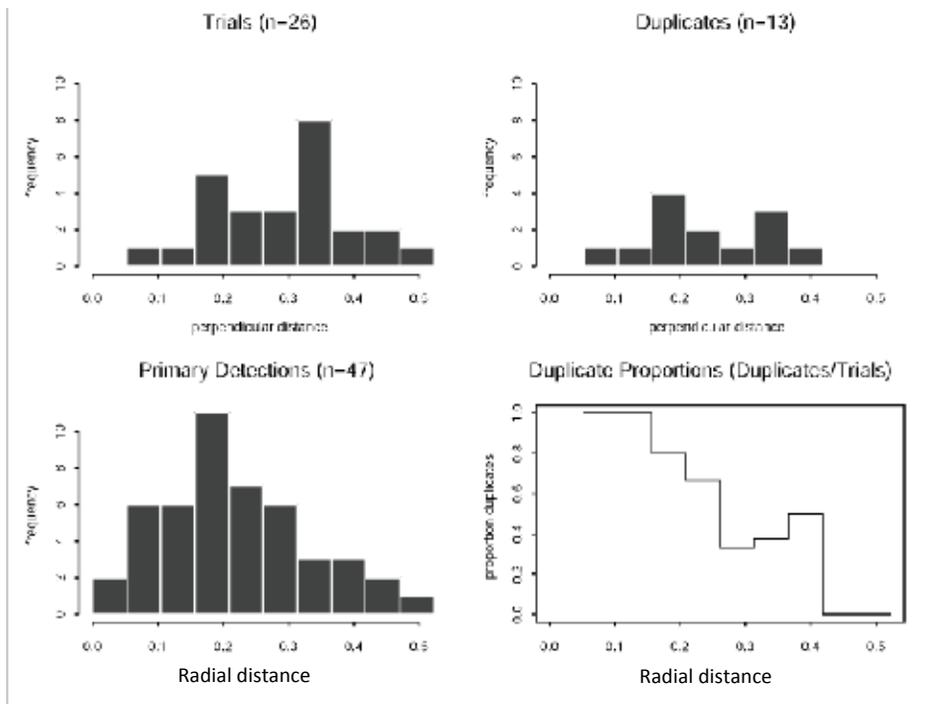
Fig. 7. *2001 Survey. Detections by observer S1 ("Trials": top left plot), observer P1 ("Primary Detections": bottom left plot), and duplicates (top right plot) for periods when S1 and P1 were paired on the right of the aircraft. The bottom right plot shows the proportions of S1 detections which were detected by P1.*

precise estimator of abundance than is obtained using all the data. A disadvantage of using data from P1 only is that abundance in stratum 7 is estimated to be zero in this case, when we know that it must be greater because P2 saw some animals there. This might lead one to prefer the corrected estimate using all observers' data.

DISCUSSION AND CONCLUSIONS

The 1987 cue–counting survey was unusual in that the magnitude of the measurement error is the largest we have encountered. It is large enough to cause substantial bias in estimated $h(0)$ and abundance if neglected. We have therefore estimated abundance incorporating a measurement error process which was estimated from pairs of distances to duplicate detections. While the estimation method we used allows bias as well as random error in distance

estimation to be incorporated, bias cannot be estimated from the duplicate distance estimates and has therefore been assumed to be zero.

One of our main objectives in re-analyzing the 1987 survey data was to resolve the discrepancies between our abundance estimates and those produced by Hiby *et al.* (1989). We were not able to reproduce the estimate of Hiby *et al.* (1989) and our most comparable estimate is 53% higher with a higher cv. However there are differences between our data and those used by Hiby *et al.* (1989) and there are also some differences between the methods used. In particular, Hiby *et al.* (1989) use data grouped into distance intervals whereas our analysis uses ungrouped distances.

Comparison of our Table 1 with Table 2 in Donovan and Gunnlaugsson (1989) reveals some fairly small differences in the amount of survey effort and the number of cues seen by

Table 4. 2001 survey. The estimated abundance and related parameters from the original strata. $\hat{h}(0)$ is the estimated slope of the pdf of radial distances at the origin. Only cues detected within 0.54 nm on the side with P1 are included; areas are in nm²; time (*T*) is in hours.

Stratum	Area	*T*	*n*	*n/T*	$\hat{h}(0)$	$\hat{N}$	%cv	95%CI
1	4,418	8.21	44	5.36		7,678	22	(4,984-11,830)
2	3,988	3.00	4	3.00		1,728	52	(631-4,730)
3	14,066	5.60	6	5.60		4,887	38	(2,343-10,195)
4	12,392	10.19	19	10.19		7,500	25	(4,641-12,120)
5	10,782	3.86	4	1.04	26.50	3,623	48	(1,436-9,157)
6	3,602	3.36	16	3.36		5,562	32	(2,955-10,467)
7	14,384	5.32	0	0.00		0	–	(—— ——)
8	3,728	3.81	23	6.02		7,292	46	(3,202-16,607)
9	18,186	5.50	5	0.91		5,360	55	(1,907-15,065)
Total	**85,546**	**48.85**	**121**	**35.48**		**43,633**	**19**	**(30,148-63,149)**

Table 5. 2001 survey. The estimated abundance and related parameters from the original strata. $\hat{h}(0)$ is the estimated slope of the pdf of radial distances at the origin. Only cues detected within 0.54 nm on either side of the aircraft are included; areas are in nm²; time (*T*) is in hours.

Stratum	Area	*T*	*n*	*n/T*	$\hat{h}(0)$	$\hat{N}$	%cv	95%CI
1	4,418	8.21	79	9.62		5,998	20	(4,038-8,910)
2	3,988	3.00	10	3.34		1,879	44	(793-4,455)
3	14,066	5.60	11	1.96		3,898	31	(2,128-7,141)
4	12,392	10.19	34	3.34		5,839	22	(3,817-8,908)
5	10,782	3.86	8	2.07	23.06	3,155	39	(1,469-6,774)
6	3,60 2	3.36	17	5.06		2,571	30	(1,416-4,666)
7	14,384	5.32	3	0.56		1,146	63	(352-3,730)
8	3,728	3.81	36	9.44		4,966	46	(2,048-12,039)
9	18,186	5.50	9	1.64		4,197	52	(1,565-11,261)
Total	**85,546**	**48.85**	**207**	**46.65**		**33,649**	**17**	**(23,988-47,200)**
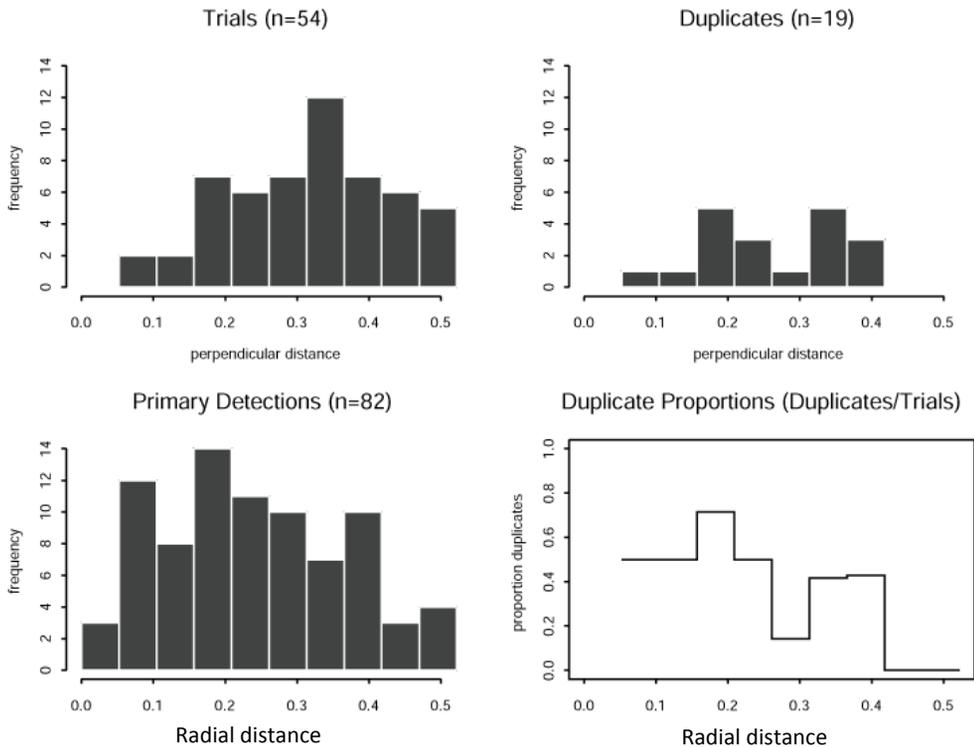
Fig. 8. *2001 Survey. Detections by observers P1 or P2 ("Trials": top left plot), observer S1 ("Primary Detections": bottom left plot), and duplicates (top right plot) for periods when S1 and either P1 or P2 were paired on the right of the aircraft. The bottom right plot shows the proportions of P1 and P2 detections which were detected by S2.*
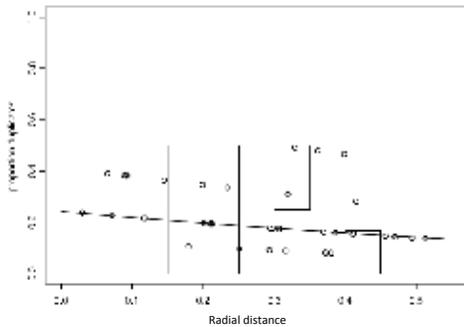
Fig. 9. *2001 Survey. Estimated logistic detection function (smooth curve) for observer P2 from periods when S1 and P2 were paired on the right of the aircraft. The smooth curve is a weighted average of separate curve for each Beaufort level, with weight proportional to the estimated number of groups available for detection at that Beaufort level. Dots are estimated detection probabilities for individual detections; the histogram shows the observed duplicate proportions.*
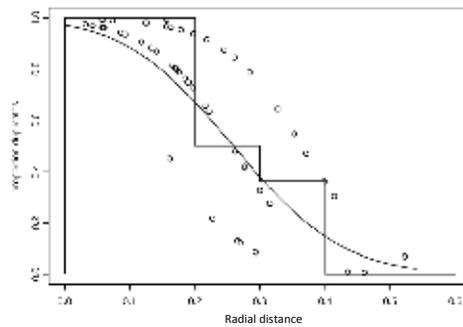
Fig. 10. *2001 Survey. Estimated detection function (smooth curve for observer P1 from periods when S1 and P1 were paired on the right of the aircraft. Dots are estimated detection probabilities for individual detections; the histogram shows the observed duplicate proportions.*
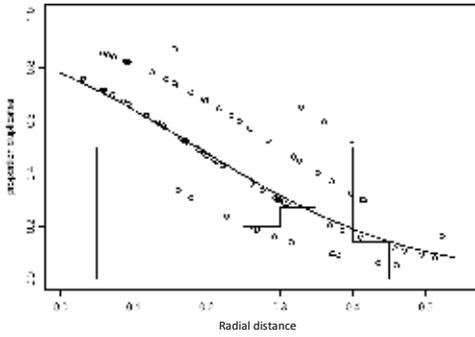
Fig. 11. *2001 Survey. Estimated detection function (smooth curve for observer P1 from periods when S1 and either P1 or P2 were paired on the right of the aircraft. Dots are estimated detection probabilities for individual detections; the histogram shows the observed duplicate proportions.*
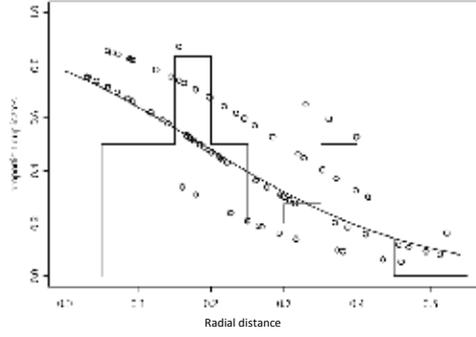
Fig. 12. *2001 Survey. Estimated detection function (smooth curve for observer P1 and P2 together from periods when S1 and either P1 or P2 were paired on the right of the aircraft. Dots are estimated detection probabilities for individual detections; the histogram shows the observed duplicate proportions.*
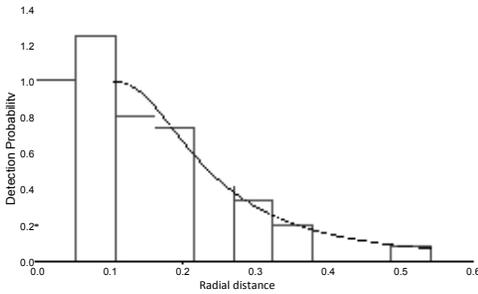
Fig. 13. *2001 Survey. Estimated detection function for all observers, assuming p(0)=1 and truncating at 0.54 nm. The histogram shows observed frequencies scaled up in inverse proportion to the radial distance at the midpoint of the histogram bar.*
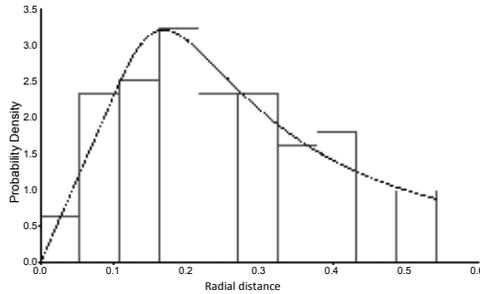
Fig. 14. *2001 Survey. Estimated probability density function of distances to detections for all observers, truncating at 0.54 nm. The histogram shows observed frequencies.*
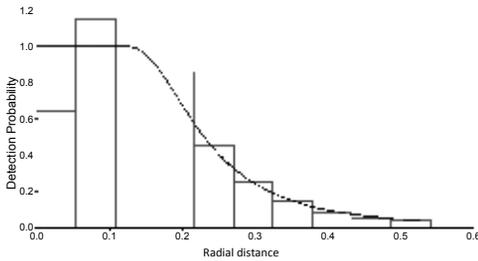
Fig. 15. *2001 Survey. Estimated detection function for observer P1 only, assuming p(0)=1 and truncating at 0.54 nm. The histogram shows observed frequencies scaled up in inverse proportion to the radial distance at the midpoint of the histogram bar.*
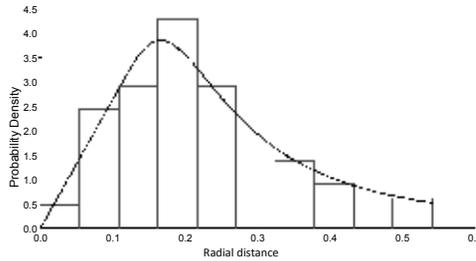
Fig. 16. *2001 Survey. Estimated probability density function of distances to detections for observer P1 only, truncating at 0.54 nm. The histogram shows observed frequencies.*

stratum. However, the main source of the difference in the abundance estimates is in the estimates of $h(0)$. Because the distance data used by Hiby *et al.* (1989) are no longer available, we are not able to say exactly what the sources of these differences are – all we have is the radial distance distribution histograms in Hiby *et al.* (1989). In Fig. 17, we have reproduced the relevant histogram from that paper and overlaid the radial distance distribution histogram of the data used in this analysis. Our data have higher frequency in the first distance interval, and this is consistent with a higher estimated $h(0)$ and higher estimated abundance. It seems therefore that a combination of differences in data and analysis methods are responsible for the differences in estimated abundances.

We have carefully checked and verified the surviving dataset and we are confident that these data are internally consistent, and are the best available. In the future, we recommend that better procedures be used to fully verify and document all NASS survey data, and that dataset should be conserved in a single consistent format to be made available to analysts. In this way we hope that future problems of data validation and analysis replication can be avoided.

Our best estimate of minke abundance from the 1987 survey, using the strata defined by Donovan and Gunnlaugsson (1989), is 16,468 animals with a 95% confidence interval of (8,287; 32,724). Using the original strata, our best estimate is 24,532 animals with 95% confidence interval (13,399; 44,916). This latter estimate can be compared to the estimate of 38,071 (95% CI 25,908; 55,945) for the same strata surveyed in 2001. This implies a population growth rate λ of 4.1% (with cv=18% and 95% confidence interval 2.9% to 6.0%) annually over the period. These estimates are based on the growth model $N_{2001} = N_{1987} \times (1 + \lambda)^{14}$. The cv and 95% confidence interval were obtained by simulating 10,000 replicates of N_{2001} and N_{1987} from lognormal distributions with means equal to 24,532 and 16,468 and cv's equal to 19% and 32%, respectively, and calculating λ for each replicate. Pike *et al.* (2009b) found a 2% rate of growth, not significantly different from 0, in relative abundance for minke whales in the same area between 1986 and 2001.

All of the area covered in these surveys falls within the CIC "Small Area" of the Central Stock Area as defined by the International Whaling Commission. Whales in this area are considered part of the Central Stock, but the stock structure of minke whales on the feeding grounds is very uncertain (NAMMCO 2004).

Virtually all Icelandic minke whaling has been carried out within the survey area. Minke whaling began in coastal Icelandic waters around 1930 and continues today. The most intensive catching period was from 1965-1969, when catches averaged 451 per year, after which catches declined to an average of 292 per year from 1980-1984 (NAMMCO 1998). Minke whale hunting ceased in Iceland in 1986 and resumed at a very low level only in 2003.
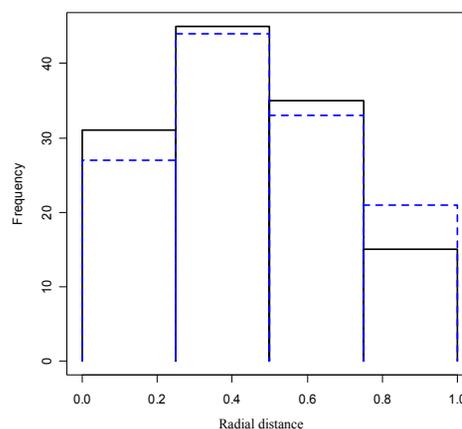
Fig. 17. *1987 Survey. Radial distance histograms of the data used by Hiby et al. (1989) (dashed line) and the data used in this paper (solid line).*

Using these estimates and associated NASS shipboard estimates from outside the aerial survey area (Pike *et al.* 2009a), the NAMMCO Scientific Committee conducted an assessment of the Central Stock in 2003 (NAMMCO 2004). The results of this assessment indicated that the Central Stock of minke whales has not been appreciably impacted by past whaling, having a current abundance of mature females that is at least 85% of the corresponding pre exploitation level. Our results suggest that the minke whale stock in the

area may be growing slowly, probably as a result of recovery from past whaling. The Trans North Atlantic Sightings Survey, to be conducted in 2007, will provide new estimates for this area using similar survey methods, so that we can determine if this positive trend is continuing.

REFERENCES

Borchers, D.L., McCracken, M. Gunnlaugsson, Th. and Burt, M.L. (MS) 1997. Estimates of minke whale abundance from the 1987 and 1995 NASS aerial surveys. Working paper SC/5/AE/2 for the NAMMCO Scientific Committee Working group on Abundance Estimates. [Available at the NAMMCO Secretariat]

Borchers, D.L., Buckland, S.T., Goedhart, P.W., Clarke, E.D. and Hedley, S.L. 1998. Horvitz-Thompson estimators for line transect surveys. *Biometrics* 54:1221-1237.

Borchers, D.L., Laake, J.L., Southwell, C. and Paxton, C.G.M. 2006. Accommodating unmodelled heterogeneity in double-observer distance sampling surveys. *Biometrics* 62:372-378.

Borchers, D.L. (MS) 2003. Analysis of NASS 1987 and 2001 minke whale cue–counting surveys taking account of distance estimation errors. Working paper SC/11/AE4 for the NAMMCO Scientific Committee Working group on Abundance Estimates. [Available at the NAMMCO Secretariat]

Borchers, D.L., Marques,T.A., Gunnlaugsson, Th. and Jupp, P. (*submitted*) Distance sampling with measurement errors.

Buckland, S.T., Anderson D.R., Burnham, K.P, Laake, J.L., Borchers, D.L. and Thomas, L. 2001. *Introduction to distance sampling*. Oxford University Press, Oxford, 432pp.

Donovan, G.P. and Gunnlaugsson, Th. 1989. North Atlantic sightings survey 1987: Report of the aerial survey off Iceland. *Rep. int. Whal. Commn*. 39:437-441.

Gunnlaugsson, Th. (1989) Report on Icelandic minke whale surfacing rate experiments in 1987. *Rep. int. Whal. Commn*. 39:435-436.

Hiby, A.R. and Hammond, P.S. 1989. Survey techniques for estimating abundance of cetaceans. *Rep. int. Whal. Commn*. (Special Issue 11):47-80.

Hiby, L., Ward. A. and Lovell, P. 1989. Analysis of the North Atlantic Sightings Survey 1987: Aerial survey results. *Rep. int. Whal. Commn*. 39:447-455.

[NAMMCO] North Atlantic Marine Mammal Commission. 2002. Report of the Ninth Meeting of the Scientific Committee. In: *NAMMCO Annual Report 2001*:147-270.

[NAMMCO] North Atlantic Marine Mammal Commission. 2004. Report of the Eleventh Meeting of the Scientific Committee. In: *NAMMCO Annual Report 2003*:135-310.

Pike, D.G., Gunnlaugsson, Th., Víkingsson, G.A., Desportes, G. and Bloch, D. 2009a. Estimates of the abundance of minke whales (*Balaenoptera acutorostrata*) from Faroese and Icelandic NASS shipboard surveys. *NAMMCO Sci. Publ*. 7:81-93.

Pike, D.G., Paxton, C.G.M., Gunnlaugsson, Th. and Víkingsson, G.A. 2009b. Trends in the distribution and abundance of cetaceans from aerial surveys in Icelandic coastal waters, 1986-2001. *NAMMCO Sci. Publ.* 7:117-142.

The R Foundation for Statistical Computing. 2008. R version 2.6.2. http://www.r-project.org/.

Thomas, L., Laake, J.L., Strindberg, S., Marques, F.F.C., Buckland, S.T., Borchers, D.L., Anderson, D.R., Burnham, K.P., Hedley, S.L., Pollard, J.H., Bishop, J.R.B. and Marques, T.A. 2005. Distance 5.0. Release Beta 5. Research Unit for Wildlife Population Assessment, University of St. Andrews, UK. http://www.ruwpa.st-and.ac.uk/distance/.