Abundance of minke whales (*Balaenoptera acutorostrata*) in the Northeast Atlantic: variability in time and space

Hans J. Skaug, Nils Øien, Tore Schweder, and Gjermund Bøthun

Abstract: Regional sighting surveys with two independent observers on each vessel were conducted each year from 1996 to 2001. Northern minke whales (*Balaenoptera acutorostrata*) are mostly solitary animals and are only available for observation at moments when they surface to breath. Thus, a stochastic point process model is developed for how the data are generated. The hazard probability of initially sighting a whale that surfaces depends on relative spatial coordinates and on other covariates. The parameters of the model are estimated by maximum likelihood. To account for interannual variation in spatial distribution of minke whales, a random effects model is developed and estimated by comparing current and past (1989 and 1995) survey data. A simulation approach is taken to remove bias from parameter estimates and to assess the uncertainty in the results. For total abundance, the result is a log-normal confidence distribution with quantiles $107\,205 \cdot \exp(0.137z)$, i.e., an abundance estimate of 107 205 with a coefficient of variation of ≈0.14. Together with these and earlier survey data, past data on catch, mark–recapture, and satellite tracking are reviewed to elucidate distribution and migration patterns in Northeastern Atlantic minke whales.

Résumé : Chaque année, de 1996 à 2001, il y a eu des inventaires visuels régionaux avec deux observateurs indépendants à bord de chaque navire. Les petits rorquals (*Balaenoptera acutorostrata*) sont généralement des animaux solitaires et ils ne sont visibles que lorsqu'ils remontent à la surface pour respirer. C'est pourquoi nous avons mis au point un modèle de processus ponctuel stochastique pour étudier comment les données sont générées. La probabilité de hasard de percevoir initialement un rorqual quand il refait surface dépend de coordonnées spatiales relatives et d'autres covariables. Les paramètres du modèle sont estimés par une méthode de vraisemblance maximale. Un modèle à effets aléatoires estimé par la comparaison des données d'inventaire actuelles et passées (1989 et 1995) permet de tenir compte de la variation d'une année à l'autre de la répartition des rorquals. L'abondance totale obtenue est une distribution de confiance log-normale avec les quantiles $107\,205 \cdot \exp(0{,}137z)$, soit une estimation d'abondance de 107 205 avec un coefficient de variation de ≈0,14. Un revue de ces données d'inventaire et de celles du passé, des données du passé sur les pêches et des données de surveillance par satellite nous a servi à élucider les patterns de répartition et de migration des petits rorquals du nord-est de l'Atlantique.

[Traduit par la Rédaction]

Introduction

Minke whales (*Balaenoptera acutorostrata*) are found throughout the North Atlantic, although their main distribution is thought to be over continental shelf areas, particularly at its edge. They undertake feeding migrations northwards in the spring and then enter areas of the Northeast Atlantic (Fig. 1). Their winter distribution in the North Atlantic is poorly known as there are few sighting records from that time of the year. Wintering minke whales in the North Atlantic have been reported from the northern coasts, from the Gulf of Mexico, and from south of Bermuda (Horwood 1990).

Exploitation of minke whales in the Northeast Atlantic by Norwegian fishermen escalated in the 1920s and reached its peak in the 1950s with a yearly catch of some 3500 animals (International Whaling Commission 1992). From 1959, the exploitation declined because of harvest regulations and came to a halt in 1987. In 1993, commercial minke whaling was started up again under Norwegian regulation based on the Revised Management Procedure developed by the IWC (International Whaling Commission). The yearly catch has been some 500 animals in the recent decennium. The abundance of minke whales feeding in the North Atlantic, and its temporal and spatial distribution, is of interest not only for whaling management, but also for fisheries management and ecological studies.

The first attempt to estimate the abundance of minke whales in the Northeast Atlantic was based on catch–effort data and yielded an estimate of about 30 000 minke whales (Ugland 1976). Based on two recaptures from 15 minke

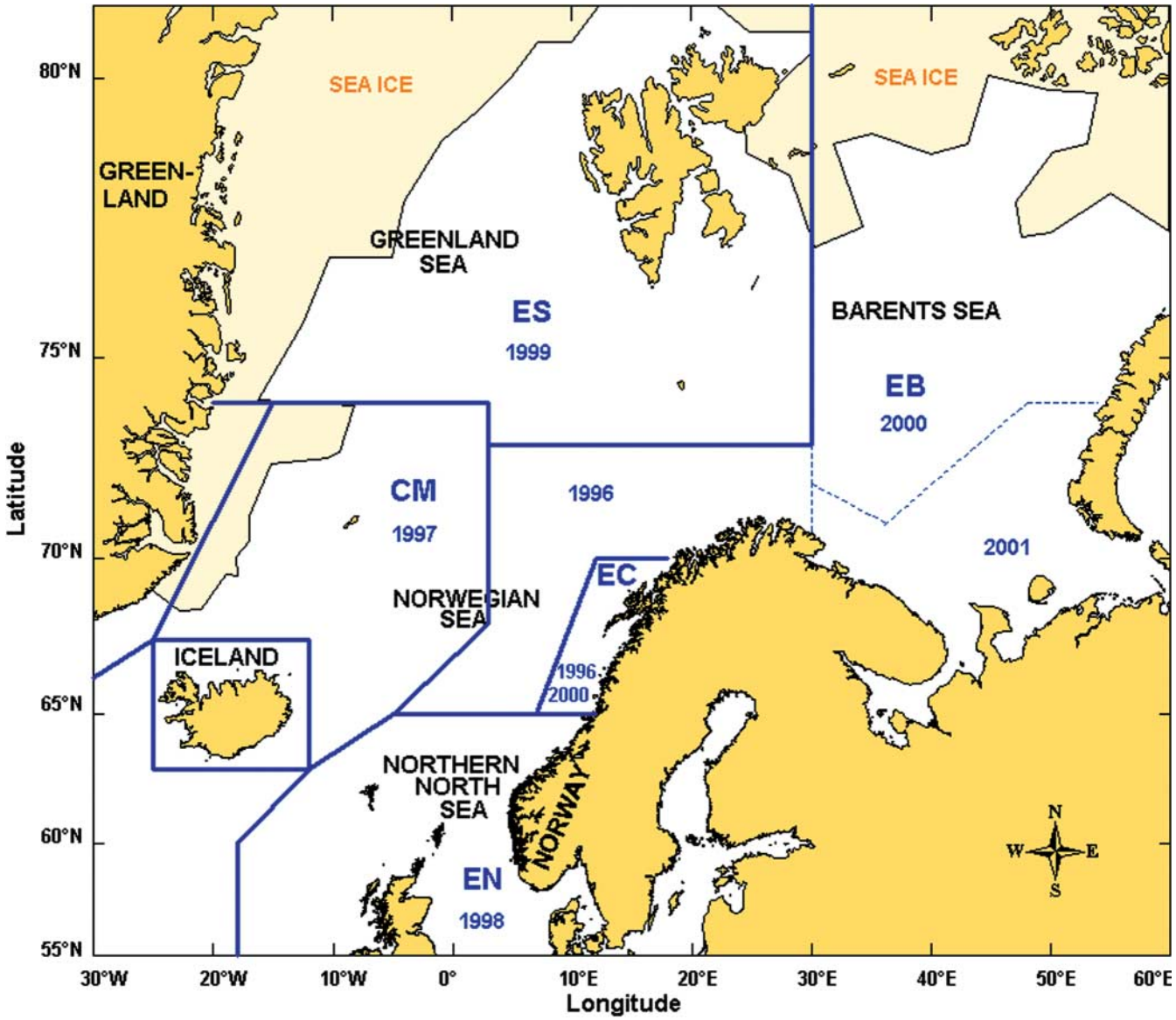
Received 31 March 2003. Accepted 24 January 2004. Published on the NRC Research Press Web site at http://cjfas.nrc.ca on 23 July 2004.
J17437

H.J. Skaug, N. Øien, and G. Bøthun. Institute of Marine Research, Box 1870 Nordnes, 5817 Bergen, Norway.
T. Schweder. Department of Economics, University of Oslo, Box 1095 Blindern, 0317 Oslo, Norway.

[1]Corresponding author (e-mail: skaug@imr.no).

 doi: 10.1139/F04-020

Fig. 1. The Northeast Atlantic split into survey regions with synoptic coverage in single years. These regions coincide with the five small areas of the International Whaling Commission (delimited by blue, solid lines): CM, EB, EC, EN, and ES, except for EB which was covered during three different years (see blue, broken, thin line). The ice edge was read from ice charts (Norwegian Meteorological Institute) for mid-July for the year when the adjacent survey blocks were covered.



whales marked in 1964 and 1965, Ugland (1976) obtained an independent estimate of 40 700 animals. Over the period 1974–1978, a total of 333 minke whales were tagged in the Barents Sea, and from recaptures of these tags, Christensen and Rørvik (1981) obtained an estimate of 113 000 animals. Of these, the number of whales recruited to the areas north of 70°N was estimated to be 55 000. Based on the same marking experiment, but with more years of recaptures, Beddington et al. (1984) found point estimates of the Northeast Atlantic stock of minke whales from 81 500 to 121 000 animals based on multiyear recaptures and 66 000 based on next-year recaptures only. The IWC concluded in 1983 that the multiyear recapture estimates should be discarded and, therefore, that a best estimate for the total stock should be 60 000 animals.

Schweder et al. (1997) presented two independent abundance estimates of Northeast Atlantic minke whales: one based on a combination of data from shipborne sighting surveys (Fig. 2) conducted in 1988 and 1989 and the other based on a double-platform shipborne survey carried out in 1995. Here, we present an abundance estimate based on a series of regional sighting surveys conducted in the period from 1996 to 2001.

The abundance estimates presented in this paper came about in the context of regulating the harvest of minke whales in the North Atlantic. The Norwegian reservation against the moratorium on commercial whaling decided by IWC in 1982 makes the Norwegian minke whaling legal under international law. In 1982, IWC also requested its Scientific Committee to develop a management procedure for commercial harvesting of baleen whales. This work was basically completed in 1991 with a proposed Revised Management Procedure (RMP). The Commission has still not implemented any scheme for commercial whaling (http://www.iwcoffice.org), but its Scientific

Fig. 2. Double-platform shipborne sighting surveys. Observer platform A is placed in the barrel as shown, and observer platform B (not shown) is placed at the wheelhouse roof. The depicted scenario involves a single whale making three surfacings, which are all detected by platform A. Platform B sees only the last surfacing. Thus, A1 is the initial sighting of the whale, whereas A3 and B1 form a duplicate sighting, as indicated by the broken line. Because of measurement error in distance and sighting angle, the recorded relative positions of A3 and B1 differ. In the statistical analysis, the second and third surfacings form Bernoulli trials for platform B. (Artist: Åsa Gan Schweder.)

Committee assures that the Norwegian management is carried out according to the RMP.

A semi-Bayesian catch limit algorithm forms the core of the RMP (International Whaling Commission 1994*a*, 1994*b*). Based on historical catch data and absolute abundance estimates of the stock in question, catch limits are calculated via a posterior distribution for the parameters of a crude production model. To guard against uncertainties in the abundance estimates that might not have been accounted for, standard errors of abundance estimates are quadrupled in these calculations. In the implementation, the ocean basin is divided into medium-sized areas and then further into smaller areas. The catch limit algorithm is applied to small areas or to medium areas with catches distributed over small areas in proportion to current abundance estimates. The latter method is called catch cascading. The RMP is robust against a wide range of uncertainties concerning stock structure and status of the stock (Cooke 1999; Butterworth and Punt 1999).

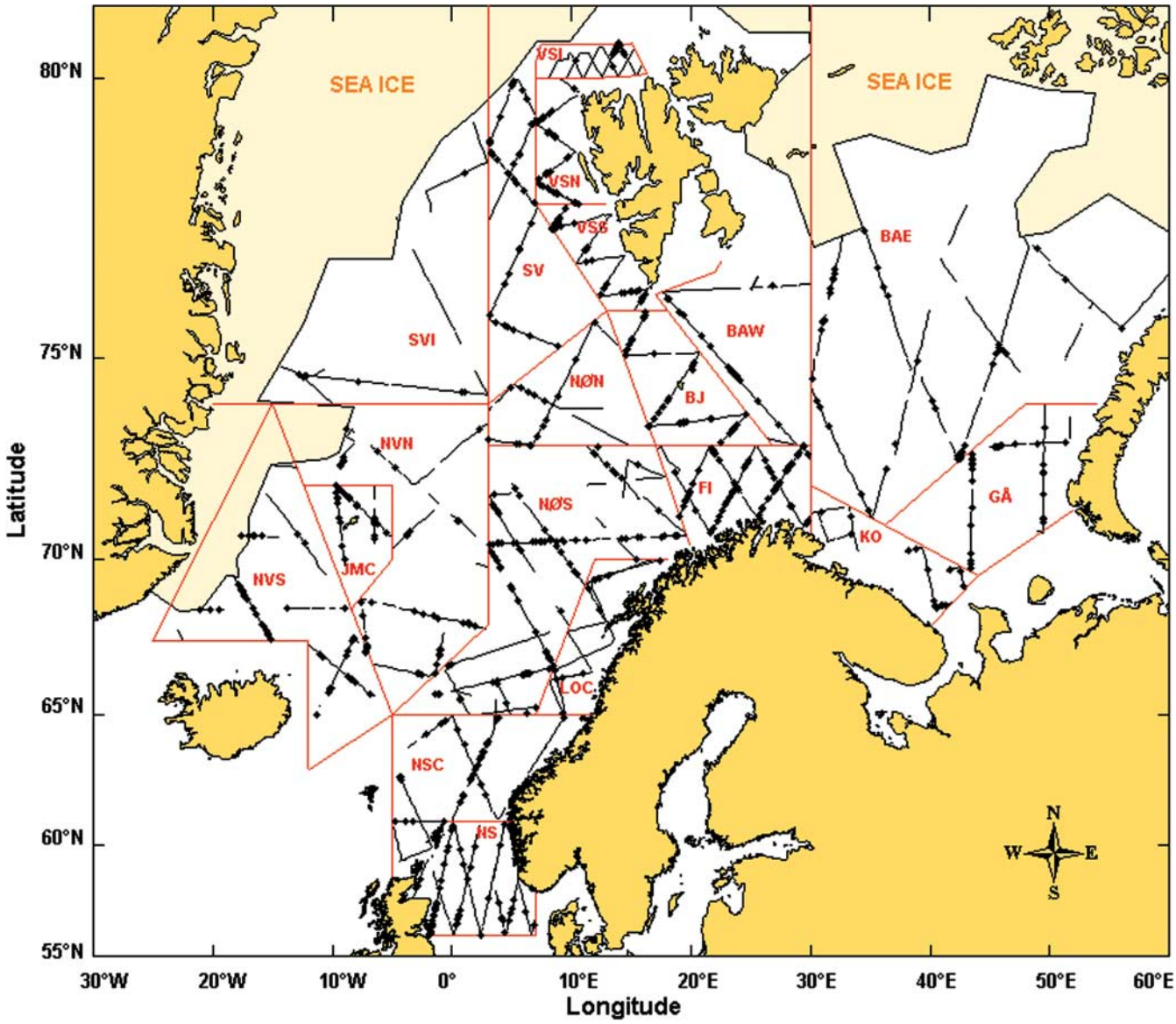
In the current RMP implementation, the North Atlantic is split into three medium areas: west, central, and east. Whaling is carried out in the small area CM (Fig. 1) of the central medium area and in the four eastern small areas (Fig. 1). For the eastern small areas, catch cascading is applied. This implementation was put in place in 1996 and was thoroughly reviewed in 2003 by the Scientific Committee. The implementation review centred on stock structure and area divisions and on the abundance estimates presented here (International Whaling Commission 2004, Annex D). The implementation was essentially upheld.

The estimates presented here are slight modifications of those in Skaug et al. (2002). These modifications result from the inclusion of additional data on surfacing rates in northern minke whales and also from some minor changes in methodology. We present an abundance estimate for each of five IWC small areas (Fig. 1), in addition to an estimate for the total survey area.

The surveys in 1996 to 2001 are visual line transect surveys with independent observer platforms. Northeast Atlantic minke whales are mostly solitary on their feeding grounds in the north (Øien 1989; Sigurjonsson et al. 1989) and are only observed in the short moments (ca. 2 s) when they surface to breath between their dives (mean diving time ca. 80 s). They are therefore regarded as showing discrete availability in the visual survey, i.e., they are observed only at discrete points in time. Following Schweder (1974, 1977), we model the sighting process as a stochastic point process in space, representing the locations of the individual whales, and in time, representing the surfacings of the whale, and a sequence of Bernoulli experiments representing whether the animal was observed or not at the given surfacing. The success probability of these Bernoulli experiments is called the hazard probability as it is the conditional probability of sighting given that the observer previously was unaware of the animal. The hazard probability depends not only on the position of the whale relative to the observer, but also on other variables such as the sea state (Beaufort), visibility, glare, etc. Buckland et al. (2001) present "standard" line transect methodology and its history, with emphasis on the detection function $g(x)$, defined as the probability of detecting an animal located at perpendicular distance x from the transect line. A basic assumption for standard line transect methods is that $g(0) = 1$, i.e., all animals located on the transect line are detected.

As the result of discrete availability, $g(0) < 1$ in visual line transecting of Northeast Atlantic minke whales. To estimate $g(0)$ and the effective strip half-width, the surveys are conducted with independent observers. The observers are placed

Fig. 3. Survey block definitions (red lines), transect lines (black lines), and minke whale (*Balaenoptera acutorostrata*) sightings (black diamonds) made while on primary search effort.



on separate platforms without audible or visual contact. Butterworth et al. (1982) were the first to realize that data from independent observers can be used to estimate effective strip half-width when $g(0) < 1$. Schweder and Høst (1992) were the first to apply hazard probability methods with independent observers in animal abundance estimation. Cooke (1997), Schweder et al. (1999), and Okamura et al. (2003), among others, have further developed the methodology.

Material and methods

Data collection 1996–2001

Surveys

During the period 1996–2001, the survey covered the Northeast Atlantic, i.e., the northern North Sea, the Norwegian Sea, the Greenland Sea, and the Barents Sea (Fig. 1). Although the North and Barents seas are shelf waters with typical average depths of about 100 m and 230 m, respectively, the Norwegian and Greenland seas are oceans with deep basins of several thousand metres. The eastern part of the Norwegian Sea was surveyed in 1996, the western Norwegian Sea in 1997, the North Sea in 1998, the Greenland Sea with adjacent shelf areas of Svalbard in 1999, and the Barents Sea in 2000 and 2001.

The total survey area was divided into blocks (see Fig. 3) based on feasibility surveys conducted in the 1980s and information extracted from catch data collected since 1938 when a licence system was introduced for Norwegian small-type whaling. Areas with assumed uniform densities were defined, also taking into consideration topographical and oceanographic features. The survey effort available within a specific block was divided between two transects to ensure at least one full coverage. Transects were constructed as zigzag tracks with a random starting point (Fig. 3). Realized track lengths in different survey blocks are provided (Table 1).

Table 1. Summary of survey results by survey block: survey region, survey block, area of survey block, survey year, transect length (L), total number of sightings ($n^{(A)} + n^{(B)}$), average total effective strip half-width ($w^{(A)} + w^{(B)}$), abundance estimate ($\hat{N}$) with associated estimate of standard deviation (SD).

Region	Block	Area (km^2)	Year	L (km)	$n^{(A)} + n^{(B)}$	$w^{(A)} + w^{(B)}$	$\hat{N}$	SD
CM	JMC	67 272	1997	615.77	38	468.35	4 432.05	921.23
	NVN	324 808	1997	1945.58	59	515.51	9 553.52	1789.07
	NVS	299 523	1997	1741.72	84	567.27	12 732.47	3426.41
EB	BAE	492 171	2000	3349.07	73	462.20	11 605.35	4887.62
	FI	94 145	1996	1658.91	120	503.59	6 761.67	1563.04
	GA	159 224	2001	1046.90	56	427.11	9 970.62	3730.05
	KO	94 034	2001	857.68	25	556.97	2 460.57	819.23
	NOS	394 260	1996	4167.66	125	453.52	13 036.93	2477.94
EC	LOC	93 839	1996	918.02	2	447.87	228.23	796.69
	LOC	93 839	2000	1074.67	10	492.18	887.06	860.98
EN	NS	259 502	1998	4032.07	146	401.10	11 713.39	3455.07
	NSC	308 918	1998	2560.35	43	419.65	6 181.51	1368.45
ES	BAW	123 082	1999	857.76	30	688.04	3 128.29	1516.23
	BJ	73 909	1999	907.22	43	917.53	1 908.97	403.40
	NON	88 141	1999	957.01	24	428.54	2 579.02	703.61
	SV	91 523	1999	967.74	59	593.68	4 699.40	1213.78
	SVI	189 072	1999	1130.84	10	432.69	1 932.07	1314.58
	VSI	8 665	1999	399.97	25	1193.38	226.93	140.33
	VSN	19 618	1999	466.39	43	587.25	1 540.04	304.38
	VSS	29 113	1999	710.94	63	597.52	2 158.87	859.82

The speed of the vessel was logged on a regular basis, as well as data on sighting conditions. The intended speed was 10 knots. In addition to the ordinary survey activity, distance and angle estimation experiments were conducted. In these experiments, the survey vessel was operated in the ordinary modus and the observers were instructed to estimate distance and angle to a stationary buoy, the exact position of which could be determined by the vessel's radar. Further details about the regional surveys are given by Skaug et al. (2002).

Data on surfacing rates for minke whales have been collected by VHF radio-tagging. The data used here are the same as those used in previous studies (Schweder et al. 1997). In addition, surfacing data from five minke whales were collected in 2001 and 2002. Most of the tagged whales were sampled in near-coastal waters within IWC small areas EC and ES. In addition, there is one whale from the North Sea (EN) and one whale from the Norwegian Sea (EB).

Observational routines

Observations of whales were made by naked eye from two platforms: platform A (a barrel placed 15 m above sea level; Fig. 2) and platform B (the wheelhouse roof), each manned by a team consisting of two observers. Each vessel had four such observer teams working 2-h shifts. With a few exceptions, the composition of teams did not change within a specific year. Most observers had experience from earlier surveys, and most had participated in all years (1996–2001). Within a team, one of the observers was instructed to scan the port 45° sector, while the other scanned the starboard 45° sector. Observers had side view, and sightings were also made outside the protocol sector. We retained sightings within the full 180° sector ahead.

The unit of observation was a track of observed surfacings, with estimated time and relative position (radial distance r and sighting angle θ (the angle between the sighting line and the track line)) recorded for every sighted surfacing. Radial distances were estimated visually, whereas sighting angles were measured using an angle board. Time was measured automatically when a button was pushed to allow the relative position to be recorded on a tape. A track represents the sighted surfacings judged by the observer to belong to the same whale. For a certain proportion of observations, the position was incompletely recorded, i.e., either r or θ was missing. This proportion was much lower for the initial sightings than for resightings. Because distance estimation is uncertain at long distances, only observations for which $r \leq 2000$ m were included in the analysis below. Further, because detection probabilities at short distances may vary substantially across platforms, observations for which $r \leq 100$ m were left out.

In addition to information directly related to sightings of whales, data on environmental conditions (weather conditions, visibility, glare, and Beaufort state; Table 2) were recorded. Certain levels of these covariates were combined to obtain a more parsimonious model (Table 2). Individual observers were grouped into three categories according to their ability to detect whales at long distances. This classification was based on the general impression that the cruise leaders had gained during the surveys and as little as possible on the actual data. On this basis, every combination of observers into teams that occurred in the survey was classified as either "long" or "short" according to their assumed ability to detect whales at long distance. In the analysis below, these two team categories were assigned a separate ρ_r parameter

Table 2. Covariates recorded on an hourly basis during the survey.

		Transformed covariate[a]		
Covariate	Description	Abbreviation	Levels	Definition
Platform	Platform indicator	P	A, B	
Weather	12 categories	W	W0, W1	W0, clear sky; W1, cloudy
Visibility	Numerical (metres)	V	High, low	High, larger than 3000 m
Glare	4 categories	G	G0, G1	G0, no glare; G1, glare
Beaufort	0–12 scale[b]	B	BI, BII, BIII	BI, 0–1; BII, 2; BIII, 3–4
Observer/team	Individual observer codes	T	Short, long	See Material and methods

[a]Transformed covariate indicates the aggregation of covariate levels used in the analysis.
[b]The survey effort used in the analysis only involves Beaufort states 0–4.

(see eq. 7 below). Our a priori belief was that $\rho_r^{\text{long}} > \rho_r^{\text{short}}$, although this constraint was not put on the model during parameter estimation. The proportion of realized effort time (sum over both platforms) with long teams was 60%.

Two teams were concurrently on watch. Neither team could be seen or heard by the other. Each team reported to the cruise leader but was not informed of sightings made by the other. The two teams were thus independent.

Statistical methodology 1996–2001

The first step in the analysis was to match tracks from the two platforms corresponding to the same whale. This matching was performed by an automatic routine mainly acting on recorded times. The parameters of the hazard probability function were then estimated by maximizing the likelihood of the observed data, disregarding matching errors and other complications. A bias-correction procedure was applied to the resulting parameter estimates to account for matching errors, measurement error in r and θ, etc. Finally, the effective search area was calculated from the estimated model, and abundance estimates were obtained by equating whale density in the whole survey stratum with observed whale density over the estimated effective search area.

The bias correction of parameter estimates was conducted by simulating the survey under realistic assumptions and for a sequence of parameter vectors. Simulations also provided standard errors on parameter estimates. Bias-corrected abundance estimates and coefficients of variation for each of the yearly surveys were also obtained from simulations, as were confidence distributions of abundance representing confidence intervals at arbitrary levels.

Finally, a combined abundance estimate for the Northeast Atlantic stock of minke whales was obtained from the sequence of yearly regional surveys. In addition to the sampling variability inherent in the regional surveys, there was a component of variation resulting from varying spatial minke whale distribution over the years 1996–2001, when the regional abundance estimates were combined to an abundance estimate for the entire stock. This additional variance was obtained by fitting a random effects model representing the changes in spatial distribution of the stock over the survey area and period.

Matching tracks

The likelihood function developed below was based on matched pairs of tracks supposedly representing the same whale seen from the two platforms and on single tracks without any match. Identification of matching tracks, and matching surfacings within tracks, was done by an automatic routine (Schweder et al. 1997), because it could be applied to numerous replicates of simulated survey data. Because it was based on a finite number of rules on observer behaviour, the automatic routine occasionally made mistakes that were easy for the human eye to spot. Because the same set of rules were used in the simulations, the problem did not occur for the simulated data. Thus, it was found necessary to augment the matching results for the real data with a small number of manually judged duplicates.

Each matched pair of tracks, such as that shown in Fig. 2, as well as each track without a match, was decomposed as follows. A combined initial sighting was the first observed surfacing of the whale. It was sighted by platform A, platform B, or both platforms simultaneously. For each combined initial sighting, there was an associated trinomial trial. The outcome of the trial was denoted by u, with different outcomes $u = A$ (seen from platform A only), $u = B$ (seen from platform B only), or $u = AB$ (seen from both platforms simultaneously).

Situations in which one platforms detected the whale before the other ($u = A$ or $u = B$) provided additional information on the conditional detection probability. Assume for simplicity that platform A detects the whale before platform B (Fig. 2), then each subsequent surfacing detected by A sets up a Bernoulli trial with outcomes seen or not seen by platform B. The initial sighting in the A track does not count as a Bernoulli trial as it is included as a trinomial trial. Trials were only registered until the first success for platform B occurs.

Hazard probability model

The starting point for the standard line transect theory (Buckland et al. 2001) is the detection function $g(x)$, defined as the probability that an animal located at perpendicular distance x from the transect line is detected. From an estimate of $g(x)$, one can obtain an estimate of the effective strip half-width:

$$(1) \qquad w = \int_0^\infty g(x)\,\mathrm{d}x$$

which is the key parameter in the abundance estimation. For animals with discrete availability, it is possible to start the modelling at a more fundamental level, that is, we can model the probability of detecting individual surfacings.

Let the hazard probability function $Q(r, \theta)$ be defined as the probability of sighting a whale that makes a surfacing at position (r, θ), given that the observer is previously unaware

of the whale. The other component of the hazard probability model is the stochastic point process governing the availability of individual whales for detection, i.e., the surfacing process. As shown below, the surfacing process together with Q determine $g(x)$, and hence w, as well as the probability density for the position of the initial sighting of a whale.

Consider first a single observer platform with hazard probability function Q. The bivariate probability density of the position of an initial sighting is denoted by $f(r, \theta)$. It is simpler mathematically to work with Cartesian coordinates $(x, y) = (r\sin(\theta), r\cos(\theta))$ than with polar coordinates (r, θ). Here, y is the forward distance along the transect line. Under the assumption that dive times follow a Poisson point process with intensity α, it may be shown (Schweder et al. 1996) that

$$(2) \qquad f(x, y) = w^{-1}\frac{\alpha}{v}Q(r_{x,y}, \theta_{x,y})\exp\left\{-\frac{\alpha}{v}\int_y^\infty Q(r_{x,u}, \theta_{x,u})\,du\right\}$$

where v is the vessel speed, $r_{x,y} = (x^2 + y^2)^{1/2}$, and $\theta_{x,y} = \arctan(x/y)$. Because

$$g(x) = w\int f(x,y)dy$$

(Buckland et al. 2001, p. 53), it follows that the detection function is given as

$$(3) \qquad g(x) = 1 - \exp\left\{-\frac{\alpha}{v}\int_0^\infty Q(r_{x,y}, \theta_{x,y})\,dy\right\}$$

When there are two independent observer platforms, A and B, with separate hazard probability functions Q^A and Q^B, then the hazard probability of the combined platform $A \cup B$ is

$$(4) \qquad Q^{A\cup B}(r, \theta) = 1 - (1 - Q^A(r, \theta))(1 - Q^B(r, \theta))$$

Each whale sighted by $A \cup B$ sets up an experiment with trinomial outcome $u \in \{A, B, AB\}$, as explained above. Conditionally on the position (r, θ), the probability distribution of u is

$$(5) \qquad q(u) = \{Q^{A\cup B}(r,\theta)\}^{-1}\cdot\begin{cases} Q^A(r,\theta)\{1 - Q^B(r,\theta)\}, & u = A \\ Q^B(r,\theta)\{1 - Q^A(r,\theta)\}, & u = B \\ Q^A(r,\theta)Q^B(r,\theta)\}, & u = AB \end{cases}$$

Assume that Q belongs to a parametric class of hazard probability functions. The contribution to the likelihood function coming from a matched pair of tracks (or a track without a match) may be decomposed as follows: the probability density (eq. 2) of the position of the initial sighting, the probability (eq. 5) in the trinomial experiment, and subsequent Bernoulli trials in situations where the initial sighting is made only by one platform ($u = A$ or $u = B$). These likelihood components are conditionally independent and hence are multiplied together. The likelihoods for different matched pairs or unmatched tracks are also independent.

A 15-point Gauss-Laguerre quadrature formula (Press et al. 1992, p. 151) is used to evaluate the integrals involved in eqs. 1, 2, and 3. The likelihood function is maximized using the optimization software AD Model Builder (Fournier 2001).

The following parametric class of hazard probability function is inherited from Schweder et al. (1997):

$$(6) \qquad Q(r, \theta) = \mu Q_1(r)Q_2(\theta)$$

where $Q_1(r) = h(-\lambda_r(r - \rho_r))/h(\lambda_r\rho_r)$ and $Q_2(\theta) = h(-\lambda_\theta(\theta - \rho_\theta))/h(\lambda_\theta\rho_\theta)$ with $h(x) = \exp(x)/(\exp(x) + 1)$ being the logistic function. The basic parameters λ_r, ρ_r, λ_θ, ρ_θ, μ of the model have the following interpretations: μ is the hazard probability at the origin (where the observer is placed), ρ_r is the distance (in metres) at which the hazard probability has dropped to 50%, and λ_r is the steepness of the hazard probability curve at distance ρ_r. There parameters ρ_θ and λ_θ have the same interpretation as ρ_r and λ_r, respectively, but for sighting angle instead of distance.

The basic parameters are allowed to depend on covariates (see Table 2) through exponential and logistic link functions:

$$(7) \qquad \rho_r = \exp(\eta_r)$$

$$\rho_\theta = \exp(\eta_\theta)$$

$$\mu = \exp(\eta_\mu)/\{1 + \exp(\eta_\mu)\}$$

where η_r, η_θ, and η_μ are linear predictors (linear combinations of covariate effects). To allow platforms A and B to have different hazard probability functions (Q^A and Q^B), the intercept term in the linear predictors may be platform-specific. The vector of all parameters in the model (λ_r, λ_θ, and the parameters associated with η_r, η_θ, and η_μ) is denoted by β.

Abundance estimation

For the purpose of abundance estimation, the subareas EN, EC, EN, EC, and CM are split into survey blocks (Fig. 3) believed to be homogeneous with whale density.

Changes in the values of covariates will cause w to vary along the transect line. Let the effective strip half-width at position l be denoted by $w(l)$ and the total survey length be denoted by L. The average effective strip half-width in the survey block is given as

$$(8) \qquad \overline{w} = \frac{1}{L}\int_0^L w(l)\,dl$$

The abundance estimate $\hat{N}$ for the survey block is obtained by equating whale density in the survey block with observed whale density over the effective search area:

$$(9) \qquad \hat{N} = \frac{n^{(A)} + n^{(B)}}{2L(\overline{w}^{(A)} + \overline{w}^{(B)})}\,\mathrm{AREA}$$

where $n^{(A)}$ and $n^{(B)}$ are the total number of sighted whales for platform A and B, respectively, L is total transect length, $\overline{w}^{(A)}$ and $\overline{w}^{(B)}$ are platform-specific averages of w, and AREA is the area of the survey block. Note that a common hazard probability model is fitted for all survey blocks. This causes estimates of $\overline{w}$, and hence $\hat{N}$, from different survey blocks to be correlated.

Simulation model

The presence of measurement errors in the spatial and temporal data (related to observed surfacings), matching errors, spatial clustering of whales, and other difficulties have made it impossible for us to obtain unbiased estimates and valid standard errors without recourse to simulation. Models for these various components of the data-generating process have been estimated, as detailed below, and combine to a more realistic model for the surveys. An implementation of

this model as a computer program is used to simulate artificial data in the same format as collected in the surveys.

The program simulates a virtual sighting vessel with two independent observer platforms moving through an "ocean" of predistributed whales. The virtual whales are available for detection at discrete time points determined by dive time series obtained from radio-tagged minke whales. Each surfacing that an individual makes is detected by the observers according to the hazard probability $Q(r, \theta)$. The spatial distribution of whales is determined by a Neyman-Scott clustered-point process (Hagen and Schweder 1994). The Neyman-Scott process is characterized by the following three parameters: $\gamma^{N\text{-}S}$ (the intensity of clusters), $\mu^{N\text{-}S}$ (the average number of whales per cluster), and $\rho^{N\text{-}S}$ (the cluster radius, i.e., the standard deviation in the bivariate Gaussian distribution that governs the position of whales relative to the cluster center).

In addition, we simulate measurement errors in radial distance r, sighting angle θ, and time point of observation (see below), incomplete tracking (the observer fails to report information about subsequent surfacings), and missing values in r and θ.

A model for measurement error was developed by Schweder (1997). This model has been modified and fitted to data from the survey period 1996–2001, yielding

$$r = 0.898r'\exp[N(0,1)\cdot 1.371\cdot(r')^{-0.2375}] \quad (10)$$

$$\theta = 1.057\theta' + N(0,1)\cdot 4.826\cdot\exp[0.0117\cdot\min(|\theta'|, 55)] \quad (11)$$

where r' and θ' are the true quantities, and the $N(0, 1)$ denote standard normal random variables. The following error model for time has been adopted from Schweder (1997):

$$t = t' + \max[0,\ 7 + 3.4\cdot N(0, 1)], \quad (12)$$

where t' denotes true time.

Simulation-based inference for separate surveys

Computer simulation of a parametric model for inferential purposes is often called parametric bootstrapping (Efron and Tibshirani 1993). We use simulation to remove bias from parameter estimates obtained under the hazard probability model, to estimate standard errors, and to establish a confidence distribution (Schweder and Hjort 2002) for the abundance of minke whales. The main source of bias in the maximum likelihood estimation is the measurement error added in the simulation model, but not accounted for in the pure hazard probability model. As noted earlier, measurement error in distance and angle estimates leads to errors in the track-matching procedure, which in turn affect the estimate of the hazard probability.

Schweder et al. (1999) used the simulation model of the previous section to remove bias from the Bernoulli part of the likelihood. Here, our aim is to obtain asymptotically unbiased inference under the assumptions of the simulation model. We thus apply bias correction to the full likelihood under the pure hazard probability model. This is done by correcting the maximum likelihood estimate $\tilde{\beta}$ by Δ to

$$\hat{\beta} = \tilde{\beta} - \Delta \quad (13)$$

The correction Δ is calculated iteratively. Start with $\beta_1 = \tilde{\beta}$. At the ith step, simulate a large data set D_i^* (we use 30 times the size of the real survey data) using $\beta = \beta_i$. This leads to a maximum likelihood estimate $\tilde{\beta}_i^*$ based on D_i^*. The bias estimate at this stage is $\Delta_i = \tilde{\beta}_i^* - \beta_i$. The next large data set to simulate is based on $\beta_{i+1} = \tilde{\beta} - \Delta_i$. We have found this process to converge fairly rapidly. The process leads to an asymptotic bias correction in the sense that $\hat{\beta}$ given by eq. 13 converges to the true value of β as the amount of data becomes large (Kuk 1995).

To study the sampling distribution of the abundance estimator, 1000 data sets of the same size as the observed data set are simulated at $\beta = \hat{\beta}$. The simulations are set up with effort at covariate levels as in the real survey. (The large data sets used to evaluate Δ do also have effort similarly distributed.) For each simulated data set, we first evaluate $\hat{\beta}$ through eq. 13, then calculate the effective strip half-width w, and finally calculate $\hat{N}$ from the abundance formula (eq. 9). The number of sightings $n^{(A)} + n^{(B)}$ in the abundance formula is also generated using the simulation model, but from a different (independent) simulation replica than that used to estimate $\hat{\beta}$. For the purpose of simulating the variance of $\hat{N}$, separate Neyman-Scott parameters are used for each survey block (see Table 3), whereas in evaluating the bias correction factor Δ, the same set of Neyman-Scott parameters is used for all survey blocks.

To obtain a confidence distribution for total abundance, a few additional bootstrap runs are carried out. A confidence distribution represents confidence intervals by its quantiles. The interval between the confidence quantiles at 2.5% and 97.5% is, for example, the 95% confidence interval. The confidence density provides a graphical representation of confidence intervals at all conceivable levels of confidence and might be regarded as the frequentist analogue to the Bayesian posterior density (Schweder and Hjort 2002). The bootstrap runs are carried out at various levels of abundance around $\hat{N}$. To simplify matters, we only vary the intensity of clusters in the Neyman-Scott process. The degree of clustering and the other parameters (β) are fixed at their estimated values. For each parametric bootstrap sample, a bootstrap estimate of abundance ($\hat{N}$), is obtained as the estimate (eq. 9). From the series of these, an approximate pivot $p(\hat{N}, N)$ with cumulative distribution function F is constructed. From Schweder and Hjort (2002), the approximate confidence distribution will then simply have cumulative distribution function $C(N) = F(p(\hat{N}, N))$.

Combination of yearly regional surveys

Because of interannual variations in spatial prey distribution, the proportion of whales present in the different survey regions (Fig. 1) will vary between years. For the purpose of estimating the total minke whale abundance in the Northeast Atlantic, this variation has no effect when the total area used by the stock is covered in a synoptic survey (such as is assumed for the years 1989 and 1995). The surveys in 1996–2001 all had partial coverage. The IWC small areas (Fig. 1) constitute areas with synoptic survey coverage in single years, except for EB which was covered during 3 years. Below, the index α refers to these small areas, together with the three parts of EB.

Table 3. Estimates of Neyman-Scott parameters based on Ripley's K function.

Region	Block	$\gamma^{N\text{-}S}$ (km^{-2})	$\mu^{N\text{-}S}$	$\rho^{N\text{-}S}$ (km)	κ	τ
CM	JMC		Poisson point process			2.36
	NVN	2.149×10^{-2}	1.57	1.154	2.6500	2.76
	NVS	7.093×10^{-4}	73.40	7.277	0.0812	4.41
EB	BAE	6.153×10^{-4}	41.40	1.213	1.1300	10.90
	FI	2.938×10^{-3}	23.30	1.677	0.3640	6.01
	GA	3.387×10^{-3}	22.90	0.848	1.3100	7.64
	KO	1.086×10^{-3}	26.50	7.029	0.0358	2.70
	NOS	3.484×10^{-4}	97.00	13.218	0.0390	3.44
EC	LOC96	4.133×10^{-5}	883.00	27.058	0.0662	9.86
	LOC00	4.133×10^{-5}	883.00	27.058	0.0662	11.30
EN	NS	4.252×10^{-6}	11 004.00	51.126	0.1910	57.40
	NSC	5.981×10^{-3}	3.48	1.549	0.0681	1.98
ES	BAW	9.425×10^{-4}	26.80	2.485	0.1850	5.99
	BJ		Poisson point process			2.85
	NON		Poisson point process			2.30
	SV	4.501×10^{-3}	12.00	2.130	0.1320	3.39
	SVI	4.631×10^{-4}	82.30	8.199	0.0678	4.29
	VSI	1.252×10^{-3}	25.90	2.484	0.1830	7.28
	VSN		Poisson point process			2.53
	VSS	4.631×10^{-4}	169.00	8.199	0.1390	8.79

Note: The parameter κ defined in eq. 18 is a measure of the average whale density within clusters. The overdispersion coefficient τ is defined as $\tau = \mathrm{Var}(n^{(A)} + n^{(B)})/E(n^{(A)} + n^{(B)})$, where both the variance and the expectation are calculated under the Neyman-Scott model.

To assess the level of additional variance in the combined estimate resulting from temporal–regional variability, we also use data from 1989 and 1995. Let N_y be the true abundance in year y in the total survey area (Fig. 1). We let N_{1989} and N_{1995} be free parameters in the model, but for the period $1996 \le y \le 2001$, it is assumed that the population grows exponentially, i.e.,

$$\log(N_y) = \log(N_{y-1}) + \omega \quad (14)$$

where ω is the unknown rate of growth (or decline). Denote by $p_{a,y}$ the proportion of whales present in the area a in year y, so that the number of whales present in the area is $p_{a,y}N_y$. The following random effects model for the ps is assumed:

$$p_{a,y} = \exp(\mu_a + \xi_{a,y})/c(y) \quad (15)$$

where $c(y) = \Sigma_a \exp(\mu_a + \xi_{a,y})$ is a normalizing factor ensuring that $\Sigma_a p_{a,y} = 1$ for all y. The fixed-effect parameter μ_a accounts for time-invariant differences in abundance between areas, and the random-effect parameter $\xi_{a,y}$ accounts for interannual changes in whale distribution. These random effects are assumed to be independent and normally distributed with zero mean and variance σ^2. The parameter σ will be referred to as the additional variance parameter.

Denote by $\hat{N}_{a,y}$ the abundance estimate for area a in year y. Assuming that $\hat{N}_{a,y}$ is conditionally log-normally distributed (given the true abundance that year), we have

$$\log(\hat{N}_{a,y}) = \log(N_y) + \log(p_{a,y}) + e_{a,y} \quad (16)$$

where the survey error terms $e_{a,y}$ are assumed to be zero mean normally distributed random variables. Equation 16 illustrates how the total error in $\hat{N}_{a,y}$ can be broken down into additional variance (variation in $p_{a,y}$) and survey error ($e_{a,y}$). There are three survey periods: 1989, 1995, and 1996–2001. The terms $e_{a,y}$ are correlated within survey period due to the use of a common estimated model for the effective strip half-width. The terms $p_{a,y}$ are correlated within year due to the normalization factor c in eq. 15 but are uncorrelated across years. Note that an estimate of the correlation structure of $\{e_{a,y}\}$ is available from the bootstrap method described earlier.

To obtain an estimate of σ, the method of restricted maximum likelihood (Punt et al. 1997) is used. Denote by $L(N, \omega, \sigma, \mu, \xi)$ the likelihood function based on the area estimates from 1989, 1995, 1996–2001, where μ and ξ are vectors and $N = (N_{1989}, N_{1995}, N_{1996})$. When evaluating L we use the assumption of conditional normality (eq. 16) together with the assumptions of eqs. 14 and 15. The restricted maximum likelihood estimate of σ is obtained by maximizing the marginal likelihood:

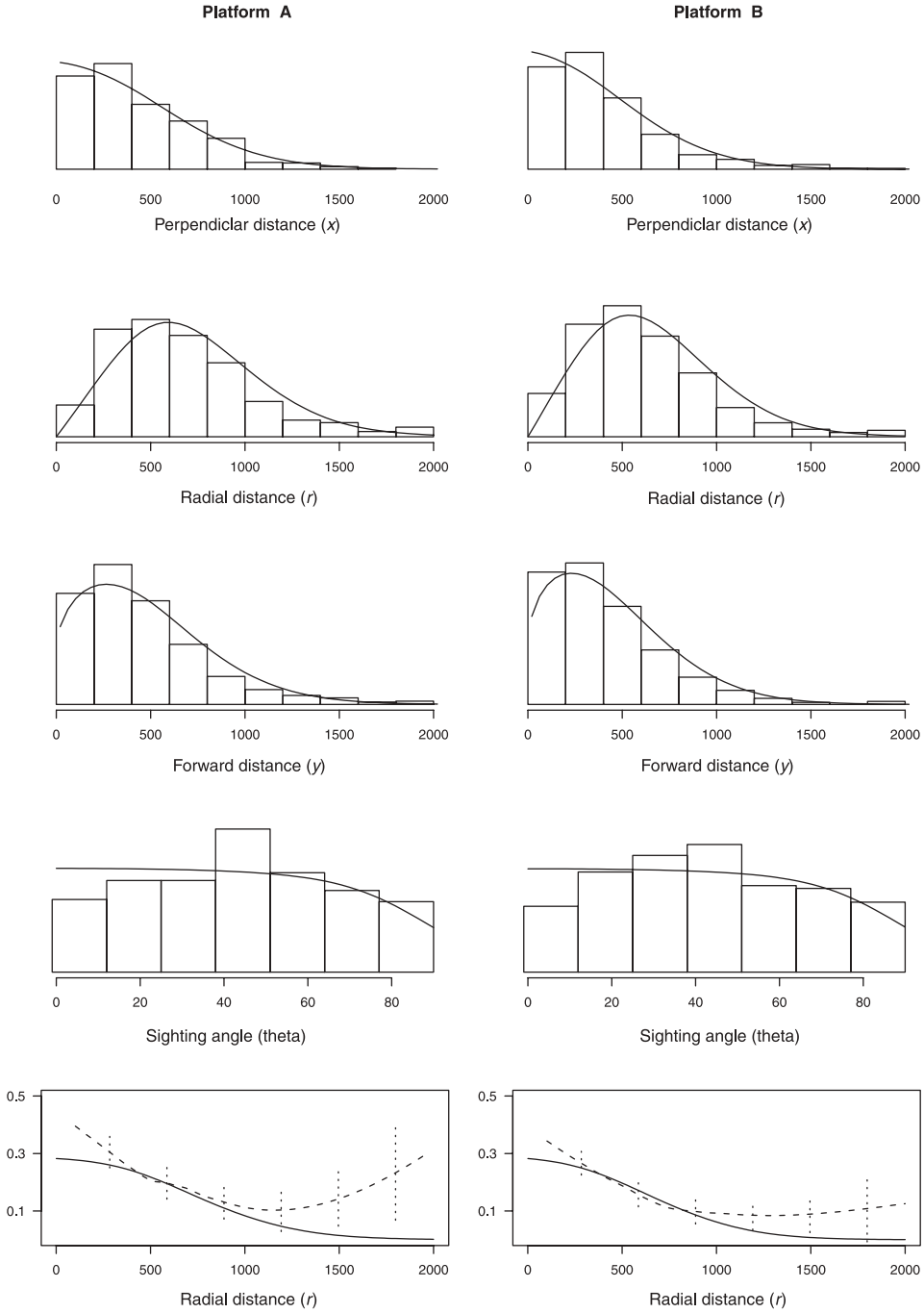
$$L(\sigma) = \int L(N, \omega, \sigma, \mu, \xi)\, dN\, d\omega\, d\mu\, d\xi \quad (17)$$

where the numerical integration is done by Laplace's method (Tierney and Kadane 1986). Because data are sparse with respect to information about σ, the resulting estimate can be expected to be biased (in addition to having a large variance). Thus, we use the simulation methodology described earlier to remove the bias in the estimate of σ. In the present setting, eqs. 14, 15, and 16 constitute the simulation model.

Results

Minke whales were found throughout the survey area, but with a spatially varying density (Fig. 3 and Table 1). A total number of 1078 sightings (totally for both platforms) was

Fig. 4. Normalized frequency distribution of observed perpendicular distances by observer platform, together with fitted probability densities (solid lines). Also shown are histograms for radial distances, forward distances, and sighting angles. The bottom two panels show estimated success probabilities by radial distance for the Bernoulli trials. The broken line is a nonparametric smoother applied to the data, and the solid line is the model predicted success probability. Vertical lines are 95% confidence intervals for the nonparametric estimate.



used in the estimation of whale density. Most of the observations (95%) were made within a strip 1000 m wide on each side of the transect line (Fig. 4). When fitting the hazard probability model, tracks with missing positional information (missing either r or θ) for all surfacings were left out. The proportion of tracks left out for this reason was approximately 1%. The track-matching routine identified 200 duplicate whales. Manual judgment identified two additional duplicate whales. The resulting number of initial sightings (for the combined platform $A \cup B$), after applying the data-truncation rules, was 870, and the number of Bernoulli trials was 623.

Likelihood analysis

Initial attempts to fit the model showed that it was not possible to simultaneously estimate the two sighting-angle parameters ρ_θ and λ_θ. As a consequence, we fixed the slope at the rather arbitrary (but not unrealistic) value $\lambda_\theta = 0.1$, which was also used by Schweder et al. (1997), and we esti-

Table 4. Comparison of different covariate models (Table 2) for the linear predictor η_r.

η_r	Deviance	df	Abundance
B+W+V+G+P+T	0	0	111 000
W+V+G+P+T	–109.5	2	109 000
B+ V+G+P+T	**–3.4**	**1**	**110 000**
B+W+ G+P+T	–8.7	1	112 000
B+W+V+ P+T	–21.3	1	111 000
B+W+V+G+ T	–16.8	1	112 000
B+W+V+G+P	–36	1	109 000
B	–86.8	5	112 000
W	–215.1	6	109 000
V	–206.7	6	104 000
G	–201.9	6	104 000
P	–199.1	6	106 000
T	–181.4	6	107 000
	–220.3	7	107 000

Note: The selected model is printed with bold type. Deviance is two times the log-likelihood ratio relative to the full model (first row), and df is the degrees of freedom to use in a likelihood ratio test. Abundance estimates (last column) are given for the total survey area without bias correction.

mated ρ_θ as a free parameter. The deviances for a set of selected covariate models are shown in Table 4, ranging from the model without covariates (bottom line) to the full covariate model (first line). A backwards model selecting scheme was used to eliminate covariates from the full model. The likelihood ratio test was used to compare nested models. "Weather" was the only covariate that could be removed from the full model (p value 0.065), leaving a model with the linear predictor $\eta_r = B + V + G + P + T$ (see Table 2 for covariate abbreviations). This model selection was carried out within the pure hazard probability model (no bias correction). A similar analysis was attempted to identify covariates affecting the level parameter μ. Here, none of the covariates yielded a significant increase in the likelihood value, and hence only the intercept term was included in the linear predictor for μ.

For the chosen model, the maximum likelihood estimate $\tilde{\beta}$ was bias-corrected using eq. 13. Parameter estimates (uncorrected and bias-corrected) are shown (Table 5), together with estimates of average effective strip half-width. Sum-to-zero constraints were imposed on the parameters associated with the different levels of categorical covariates, except for the platform covariate *P*. For instance, the table does not show the parameter estimate for level BIII of the Beaufort covariate as this parameter is determined by those of levels BI and BII. The platform effect was added only to the linear predictor of platform B. Thus, a negative estimated value of the parameter means that platform B looked at shorter range than platform A.

During the survey, 39 different combinations of covariate levels occurred. These covariate strata varied largely with respect to estimated effective strip half-width (Table 6).

Externally determined parameters

There are several parameters in the hazard probability model that were not estimated from the likelihood function. The mean surfacing rate α was estimated from 13 VHF-tagged whales to be $\alpha = 0.0129$ surfacings per second. The Neyman-Scott parameters (used in the simulation model) were estimated by matching the theoretical K function (Ripley 1977) to its empirical counterpart. A separate Neyman-Scott model was fitted to each survey block (Table 3). In four survey blocks (VSN, NON, BJ, JMC), the estimated Neyman-Scott model is consistent with a homogenous Poisson process (which is a special case of the Neyman-Scott model).

As seen from the table, the estimates for the survey block NS deviated strongly from those of the other survey blocks. A comparison of the empirical and theoretical K function indicated that the Neyman-Scott model did not fit data well for survey block NS. To resolve this problem, NS was split into two more homogenous subareas (the coast off Aberdeen and the rest of NS), and a separate estimate of τ (the coefficient of overdispersion) was obtained for each area (H.J. Skaug, unpublished data). From these estimates, an effort-weighted average $\tau = 8.96$ was calculated, and the Neyman-Scott parameters for NS were chosen to yield this level of overdispersion (under the constraint that $\gamma^{N\text{-}S}\mu^{N\text{-}S}$ was equal to the observed whale density in NS), yielding $\gamma^{N\text{-}S} = 4.34e-11$, $\mu^{N\text{-}S} = 1079$, and $\rho^{N\text{-}S} = 25\,563$. (Note that the superscript "N-S" refers to the term Neyman-Scott, and not to the survey block NS.) These estimates replace those in Table 3.

In the bias-correction procedure presented above, a common set of Neyman-Scott parameters was used for all survey blocks. This estimate was obtained by ranking the survey blocks according to their value of the quantity

$$\kappa = 0.5\mu^{N\text{-}S}(\gamma^{N\text{-}S} + (4\pi(\rho^{N\text{-}S})^2)^{-1}) \quad (18)$$

which is proportional to the error rate in the track-matching algorithm (Schweder et al. 1997). It was found, by excluding blocks where the fitted Neyman-Scott process was consistent with a Poisson process and using the median value of κ as the criterion, that a representative set of Neyman-Scott parameters was that of the survey block SV ($\gamma^{N\text{-}S} = 4.50e-09$, $\mu^{N\text{-}S} = 12.0$, and $\rho^{N\text{-}S} = 2130$).

Abundance estimates 1996–2001

Abundance estimates were calculated for each survey block (Table 1) using eq. 9. Estimates for IWC small areas were obtained by summing the contribution from the constituting survey blocks (Table 7). In the variance calculations, intrablock correlation was automatically accounted for by the parametric bootstrap approach. The table also contains estimates for the total survey area (TOTAL) and a separate estimate for the IWC eastern medium area. The latter forms the "Eastern North Atlantic medium area" in the IWC terminology. In the estimation of the additional variance parameter σ, it was necessary to exclude the survey block NVS, which was not covered in 1989 and 1995. Based on data from 1989, 1995, and 1996–2001, we obtained the estimate $\hat{\sigma} = 0.28$ (SD = 0.15), but when the bias correction explained earlier was applied, the estimate became $\hat{\sigma} = 0.22$. When ignoring additional variance (putting $\sigma = 0$), the corresponding estimates of standard error for the total area (TOTAL) was 10 821.

Parametric bootstrapping was carried out at 90%, 100%, and 110% of the total abundance estimate $\hat{N} = 107\,205$ in order to study the sampling distribution of the estimator of

Table 5. Estimated regression coefficients for the chosen hazard probability model.

	β_r	W, BI	W, BII	V, High	G, G0	T, Long	P, B	β_θ	β_μ	λ_r	$w^{(A)}$	$w^{(B)}$
$\tilde{\beta}$	6.459	0.374	0.068	0.175	0.146	0.15	–0.141	4.532	–0.565	0.0052	263	224
$\hat{\beta}$	6.729	0.332	0.050	0.177	0.121	0.129	–0.133	4.489	–0.932	0.0058	273	227
SD($\hat{\beta}$)	0.089	0.042	0.039	0.046	0.038	0.029	0.036	0.039	0.147	0.00021	17	14

Note: The parameters β_r, β_θ, and β_μ are the intercept terms in the linear predictors η_r, η_θ, and η_μ, respectively. Following the β_r column, the coefficients associated with covariates (Table 2) are given. Parameter estimates are given both with ($\hat{\beta}$) and without ($\tilde{\beta}$) bias correction. Estimates of standard deviation (SD) are given for the bias-corrected estimate. Effort weighted mean effective strip half-widths ($w^{(A)}$ and $w^{(B)}$) are also given.

Table 6. Effective strip half-width w (in metres) and $g(0)$ for selected combinations of covariate levels (Table 2).

Covariates									
Beaufort	Visibility	Glare	Team A	Team B	$w^{(A)}$	$w^{(B)}$	$g^{(A)}(0)$	$g^{(B)}(0)$	Time (%)
BIII	Low	G1	Short	Short	118	104	0.264	0.2468	0.9
BIII	Low	G0	Short	Long	153	178	0.304	0.3295	12.5
BIII	Low	G0	Long	Short	212	132	0.360	0.2807	12.9
BIII	Low	G0	Long	Long	212	178	0.360	0.3295	14.0
BII	High	G1	Short	Long	320	388	0.440	0.4801	0.1
BII	High	G0	Short	Long	465	566	0.520	0.5640	0.4
BI	High	G0	Long	Long	1078	879	0.715	0.6669	0.5

Note: The last column gives the proportion of realized survey time. The first and last rows show the lowest and highest effective strip half-widths, respectively, that occurred during the survey. Also included are the covariate settings that were dominate in terms of realized survey time.

Table 7. Abundance estimates ($\hat{N}$) by survey region (small areas in the terminology of the International Whaling Commission) in comparison with estimates for 1989 and 1995, together with coefficient of variation (CV), taken from Schweder et al. (1997).

	1989		1995		1996–2001	
Area	$\hat{N}$	CV	$\hat{N}$	CV	$\hat{N}$	CV
EB	34 712	0.203	56 330	0.136	43 835	0.15
ES	13 370	0.192	25 969	0.112	18 174	0.25
EC	2 602	0.249	2 462	0.228	584	0.26
EN	14 046	0.276	27 364	0.206	17 895	0.25
CM	2 650[a]	0.484	6 174[a]	0.357	26 718	0.14
Total	67 380	0.190	118 299	0.103	107 205	0.13
Eastern	64 730	0.192	112 125	0.104	80 487	0.15

Note: The bottom row (eastern) shows estimates for the eastern medium area of the International Whaling Commission.

[a]In 1989 and 1995, the estimates for CM did not include survey block NVS.

total abundance and to obtain a confidence distribution for this parameter. The number of bootstrap replicates were 1000, 1000, and 636, respectively. It turns out that the abundance estimator was practically normally distributed on the log scale (Fig. 5), with standard deviations 0.134, 0.145, and 0.129, respectively, and a pooled standard deviation of 0.137. These standard deviations are significantly different (p value = 0.003, Bartlett test). Because of the lack of linear pattern in the standard deviations, we base our pivot on the assumption that the standard deviation is constant on the log scale in the neighbourhood of $\hat{N}$. On the log scale, there is no significant bias in $\hat{N}$ (Fig. 5). The pivot is $p(\hat{N}, N) = (\log(N) - \log(\hat{N}))/0.137$, and F is the cumulative normal distribution function. The confidence distribution (Fig. 6) for total abundance is thus log-normal with median $\hat{N}$ and a scale parameter of 0.137. The confidence intervals for N are thus $\hat{N}[\exp(-0.137z), \exp(0.137z)]$ for appropriate normal quantiles z.

Discussion

Variation in abundance

Our abundance estimate at the stock level is lower for the period 1996–2001 than it was for 1995, but higher than for 1989. The p value for a change in abundance since 1995 is 0.18 (excluding the survey block in CM that was not covered in the 1995 survey). Is this evidence for a decline in the Northeast Atlantic stock of minke whales? Just as Schweder et al. (1997) found it difficult to ascribe the more dramatic increase in the abundance estimate from 1989 to 1995 to population growth alone, we find it difficult to ascribe the decline in estimates to a decline in the stock. One possible explanation for a decline in the eastern IWC medium area stock is that the eastern and central stocks are less separated than previously believed, at least on the feeding grounds.

Alternatively, the migration patterns of minke whales may vary because of as yet unidentified factors. According to the

Fig. 5. Normal probability plots of bootstrapped logged total abundance estimates. The panel to the left refers to simulation at $N = 0.9\hat{N}$, etc.

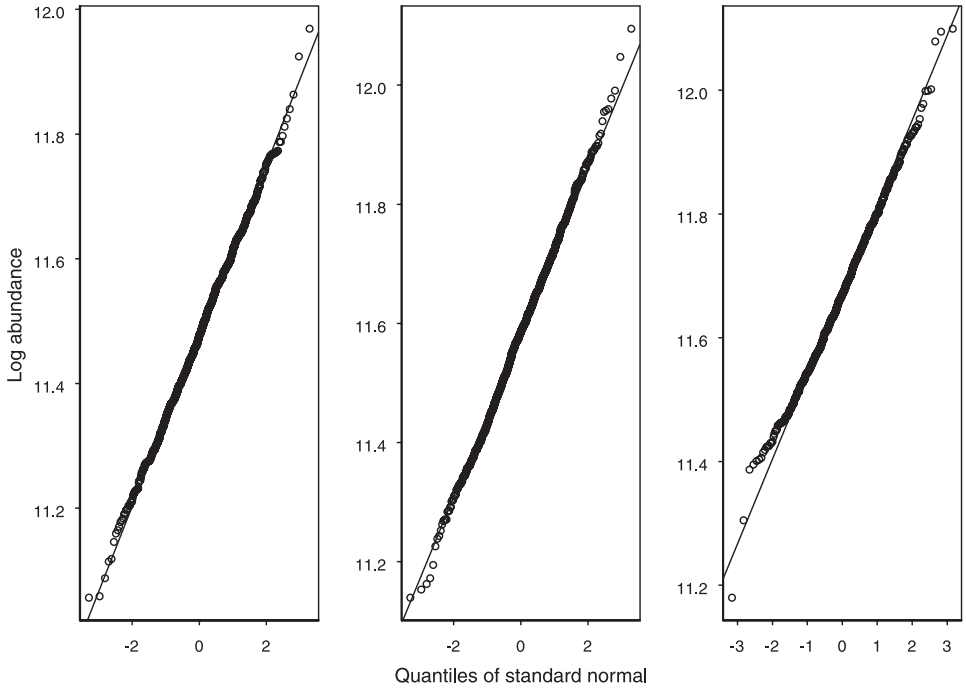
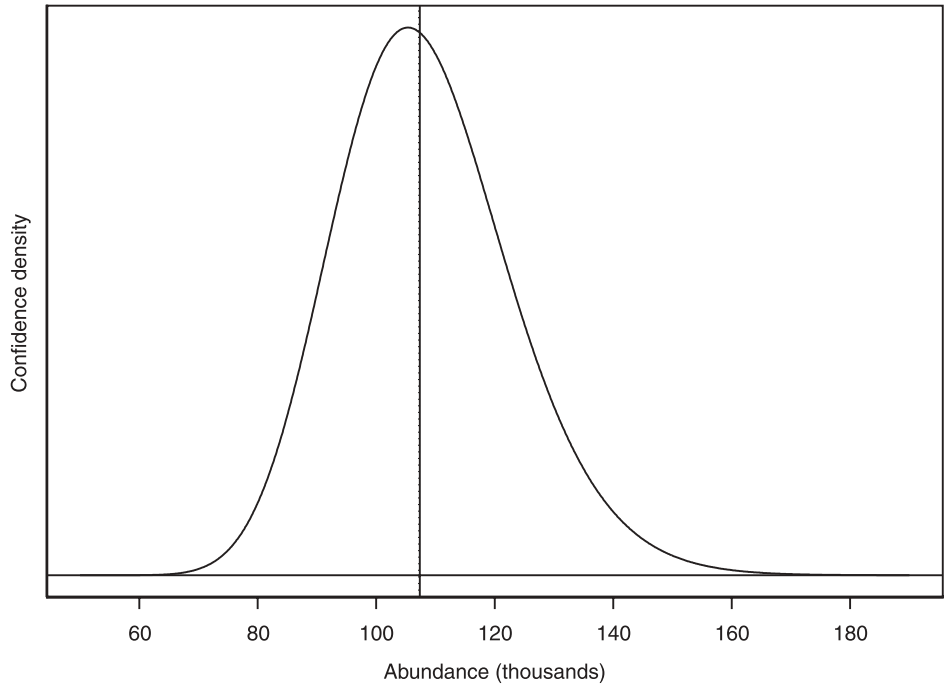


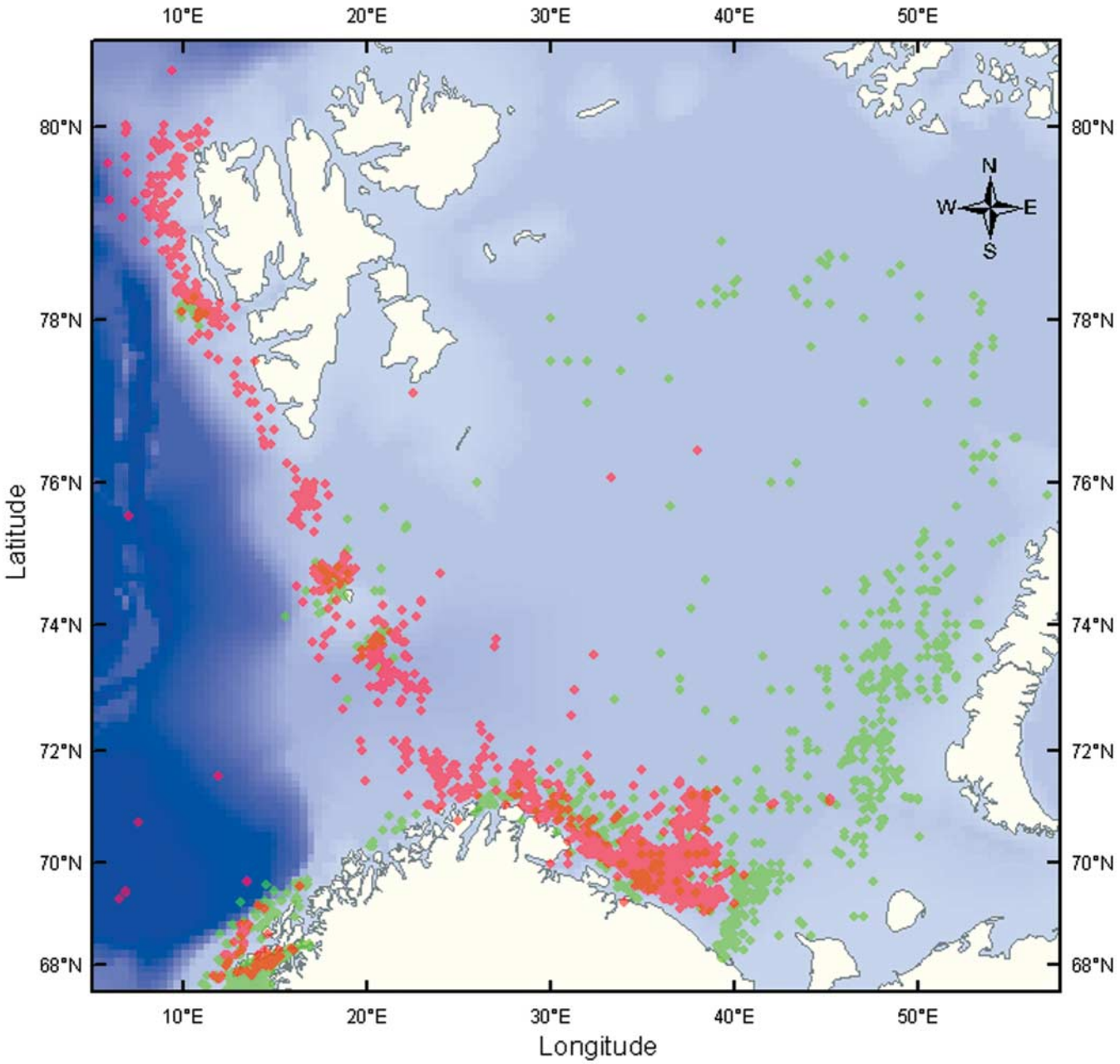
Fig. 6. Confidence density for total abundance. The median of the confidence distribution (vertical line) is nearly equal to $\hat{N}$ = 107 205.



estimated random effects model described above, the proportion of the total stock (CM and eastern small areas) present in CM each year varies randomly between 18% and 33% (lower and upper 5% prediction intervals). The increase in the estimate for the CM area from 1996 to 2001 is quite marked and could be due to more whales from the eastern areas being present in CM when it was surveyed in 1997 than was the case in 1995. Although it is possible to find alternative explanations for an apparent decline in abundance estimates from 1995 to 1996–2001 not being real, the possibility of a decline should not be ruled out. Seen over the longer period 1989–2001, however, the impression is that there might have been some growth in the population.

In 1938, a compulsory logbook system was introduced in Norwegian small-type whaling (Christensen and Øien 1990) requiring detailed information on each whale caught and all participating vessels. It is reasonable to assume that the geographical distribution of catches reflects the distribution, and changes in distribution, of whales. The catch distributions indicate that within the Barents Sea, changes have taken

Fig. 7. Geographical positions of caught minke whales (*Balaenoptera acutorostrata*) for the selected years 1952 (green) and 1980 (red). The gradient in background blue color reflects the change from several thousand metres depth in the Norwegian Sea to the shallow Barents Sea proper.



place over the years between eastern distributions (for example in 1952, see Fig. 7) and western distributions (for example in 1980, see Fig. 7).

Large temporal changes in abundance are also seen at smaller geographical scales. For instance, the survey block Kola in the Barents Sea has shown large variations in abundance during the period since 1989 (Schweder et al. 1997). The same applies to the survey block NS (North Sea), which in our survey in 1995 had a very high abundance estimate. A survey conducted in 1994 with a somewhat different methodology (Hammond et al. 2002) resulted in a much lower estimate.

Migration of minke whales into Norwegian and Arctic waters in the spring is segregated with respect to both length and sex (Jonsgård 1951; Øien 1988). From markings of minke whales, mainly in the period 1974–1978 within the Barents Sea, there is at least some indication that a regional fidelity exists (International Whaling Commission 1991). So far, instrumentation with radio and satellite tags has not given answers to questions regarding large-scale movements of minke whales in the Northeast Atlantic, but tracking over about 1-month-long periods indicates regional fidelities in Norwegian waters (Heide-Jørgensen et al. 2001), as well as Icelandic waters (Heide-Jørgensen, Greenland Institute of Natural Resources, Box 570, DK-3900 Nuuk, Greenland, unpublished data).

All of this indicates that there are large year-to-year variations in regional minke whale abundance, as is reflected in the high estimate of the additional variance parameter (σ^2).

Presumably, these variations are related to changes in abundance and distribution of possible prey species, although this has not yet been satisfactorily demonstrated. Finally, it is of course possible that a component of the minke whale stock stays outside the surveyed area and that the size of this component varies from year to year.

Estimation

The directions of the covariate effects were as expected. For the Beaufort covariate, the effective strip half-width was highest for Beaufort state "0–1" (covariate level BI), whereas it was lowest for Beaufort state "3–4" (covariate level BIII). For the team covariate, the effective strip half-width for the long category was higher than for the short category, illustrating that the classification of teams done before the analysis was meaningful. Also, the effective strip half-width was higher for the barrel (as defined previously) than for the wheelhouse roof, in agreement with the findings in Schweder et al. (1997) and with prior belief.

The bias-corrected abundance estimate for the total area is 107 205, whereas the corresponding uncorrected estimate is 110 000. Thus, the bias correction has little effect on the total abundance estimate. This could be due to cancellation of biases of different signs in the uncorrected estimate. For instance, we deliberately tuned the automatic track-matching routine (on simulated data) so that the number of duplicate pairs missed approximately equaled the number of erroneously judged duplicate pairs. The bias resulting from the manual judgement of two duplicate tracks should be small.

The measurement error models (eqs. 10 and 11) were estimated from the data collected in the distance and sighting-angle experiments conducted during the survey. For radial distance (r), it was concluded that the variability in the experimental data was representative of the variability in the real survey data. Schweder (1997), however, recommended that the variability should be estimated from the survey data. For sighting angle (θ), it was found that a higher variability was observed in the survey data than was seen in the buoy experiments. The chosen measurement error model captures this higher level of variability. The model for the difference between true and reported surfacing times (eq. 12) was adopted from Schweder et al. (1997) without modification. A priori, one expects more accurate time points in the present data than in the 1995 data, because the observers on average were more experienced than those participating in 1995.

Diagnostic plots illustrate that the distribution of distances to initial sightings are reasonably well predicted by the model, whereas the angular distribution is less well fitted. In these plots, the probability density is a weighted average (according to the number of observations) of the probability densities for the different covariate strata. We do not understand the large number of observations around 45° and the low number of sightings close to the track line. For large values of θ, the fit is reasonably good, so there is no indication that the choice $\lambda_\theta = 0.1$ made earlier is inconsistent with data. The bottom row represents diagnostic plots based on the Bernoulli trials. There are surprisingly many successes at large radial distance, especially for platform A, and more than predicted by the model. This might be explained by unmodelled heterogeneity in cue strength, partially resulting from the angle at which the whale shows its back when surfacing.

Potential sources of bias

The transect design was nonadaptive. Each detected animal generated an observational record when passed, without the vessel slowing down or deviating from the track line. Transects were planned to give a uniform coverage and not to correlate with expected whale density within survey blocks. To make full use of the available effort, two uniform layers of tracks were planned. The first layer was mandatory, and the second layer was to be covered as extensively as possible, subject to weather conditions and practicalities. The second layer was only incompletely covered. Selectivity of tracks in the second layer could lead to bias in the coverage, as also would temptation to move to regions of expected high whale density. However, the protocol specified that secondary tracks should be selected exclusively on logistic and weather considerations and that they should be discussed with the survey leader. Because weather and Beaufort Sea state is accommodated as a covariate in the analysis model, no bias should thus follow from the lack of uniform coverage that resulted from the selection of secondary tracks.

In the hazard probability method, an estimate of surfacing rate (α), obtained from radio-tagged animals, is used. A negative bias in the surfacing rate will yield a positive bias in the abundance estimates. Although our estimate of α is based on data from a limited number of individuals (13), it is not likely that the bias in our estimate of α is large. The effect of including data from five new individuals was not large (α = 0.0129 in the current paper, whereas Schweder et al. (1997) used the estimate $\alpha = 0.0130$, which was based on data from only eight of the individuals). Visual dive time data, as those of Stockin et al. (2001), have not been used in the present analysis because it is difficult to ascertain that it is a single whale that is observed without radio-marking the individual.

In the analysis, whales are assumed to be immobile except for diving. Random motion between surfacings has been found to bias abundance estimates positively when not accounted for (Schweder et al. 1996). Responsive behavior could bias abundance estimates even more. From an analysis of recorded swim directions in the 1995 data, Palka and Hammond (2001) found evidence for Northeast Atlantic minke whales avoiding the vessel when the radial distance is less than approximately 700 m. Based on the evidence available in 1995, Schweder et al. (1997) concluded that the direction and size of a potential bias resulting from random or responsive whale behavior is unknown, but if the bias is positive, its size is likely to be small in the abundance estimates.

Tracking of whales is assumed to be unbiased in the sense that the number of tracks recorded from one platform is an unbiased estimate of the number of whales actually seen. Because it is possible to break up tracks (count a whale twice) or to merge different tracks into one recorded track (count two whales as one), unbiased tracking is not self-evident. Incorrect tracking can be detected in cases where the whale has been seen by both platforms, although it is usually difficult to determine in retrospect which platform made the mistake. However, during the manual validation of the duplicate pairs, only a few such instances were discovered.

No attempt has been made to incorporate the inherent uncertainty in model selection into the estimates of variance for abundance. The variance estimates are calculated conditionally on the selected covariate model being the true model and do not reflect the fact that other covariate models might have been chosen. However, as long as the Beaufort covariate is included in the model, the abundance estimate is relatively insensitive to model choice. Further, the additional variance parameter σ was weakly determined (CV approximately 0.5 for uncorrected estimate). If σ has been underestimated, the variance estimate for the total abundance estimate will be too low.

In conclusion, the abundance estimate for the survey period 1996–2001 is practically unbiased or possibly negatively biased. In 1995 and 1988–1989, the surveys were synoptic with nearly complete coverage. In 1996–2001, the survey design could be characterized as a sequence of partial designs. Except for this difference, the survey methodology and the method of data analysis were similar for the three surveys. Differences between abundance estimates for 1989, 1995, and 1996–2001 should thus be approximately unbiased. These differences seem to indicate that the stock has increased from 1989 to 1995 and then decreased. The estimated decline since 1995 is larger than the cumulated catch (approximately 3000 minke whales) over the period. This picture of stock abundance over time prevails even when taking the standard errors in the abundance estimates into account, at least for the increase in the first period. Such variability in true stock abundance is, however, inconsistent with conventional population dynamics models for minke whales (International Whaling Commission 1991) and is likely to be less than estimated. A potential explanation is that the fraction of the stock that is present in the survey area varies from year to year, either because there is substantial contact between the central and the eastern stocks or because a variable stock component keeps to more southern latitudes at a varying degree.

Acknowledgments

This work has been generously funded by the Norwegian Government and the Norwegian Research Council (grant No. 111043/120). We are grateful to Kjell Arne Fagerheim and Siri Hartvedt at the Institute of Marine Research for preparing the data. We are also grateful to Magne Aldrin, Marit Holden, Ingunn Friede Tvete, and Gro Hagen at the Norwegian Computing Center for performing the data analysis leading to the models of eqs. 10 and 11 and to Table 3 (see Skaug et al. (2002) for further details). We also thank all of the team leaders and observers who participated in the surveys during the 6-year period. Finally, our active discussion colleagues in the Scientific Committee of the IWC are gratefully acknowledged, as well as two anonymous referees.

References

Beddington, J.R., Cooke, J.G., Christensen, I., Øritsland, T., Øien, N., and Rørvik, C.J. 1984. Assessments of the Northeast Atlantic stock of minke whales. Rep. Int. Whal. Comm. **34**: 285–91.

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., and Thomas, L. 2001. Introduction to distance sampling: estimating abundance of biological populations. Oxford University Press, Oxford, UK.

Butterworth, D.S., and Punt, A.E. 1999. Experiences in the evaluation and implementation of management procedures. ICES J. Mar. Sci. **56**: 985–998.

Butterworth, D.S., Best, P.B., and Basson, M. 1982. Results of analysis of sighting experiments carried out during the 1980/81 southern hemisphere minke whale assessment cruise. Rep. Int. Whal. Comm. **32**: 819–834.

Christensen, I., and Øien, N. 1990. Operational patterns of the Norwegian minke whale fishery. Rep. Int. Whal. Comm. **40**: 343–347.

Christensen, I., and Rørvik, C. 1981. Analysis of markings and recaptures of minke whales in the Barents Sea 1974–79. Rep. Int. Whal. Comm. **31**: 255–257.

Cooke, J.G. 1997. An implementation of a surfacing-based approach to abundance estimation of minke whale from shipborne surveys. Rep. Int. Whal. Comm. **47**: 513–528.

Cooke, J.G. 1999. Improvement of fishery-management advice through simulation testing of harvest algorithms. ICES J. Mar. Sci. **56**: 797–810.

Efron, B., and Tibshirani, R. 1993. An introduction to the bootstrap. Chapman and Hall, New York.

Fournier, D. 2001. An introduction to AD MODEL BUILDER version 6.0.2 for use in nonlinear modeling and statistics. (Available from http://otter-rsch.com/admodel.htm.)

Hagen, G., and Schweder, T. 1994. Point clustering of minke whales in the Northeastern Atlantic. *In* Developments in marine biology. Vol. 4. Whales, seals, fish and man. *Edited by* A.S. Blix, L. Walløe, and Ø. Ulltang. Elsevier Science BV, Amsterdam. pp. 27–34.

Hammond, P., Berggren, P., Benke, H., Borchers, D., Collet, A., Heide-Jørgensen, M., Heimlich, S., Hiby, A., Leopold, M., and Øien, N. 2002. Abundance of harbour porpoise and other cetaceans in the north sea and adjacent waters. J. Appl. Ecol. **39**: 361–376.

Heide-Jørgensen, M., Nordøy, E., Øien, N., Folkow, L., Kleivane, L., Blix, A., Jensen, M., and Laidre, K. 2001. Satellite tracking of minke whales (*Balaenoptera acutorostrata*) off the coast of northern Norway. J. Cetacean Res. Manag. **3**(2): 175–178.

Horwood, J. 1990. Biology and exploitation of the minke whale. CRC Press, Inc., Boca Raton, Fla.

International Whaling Commission (1991). Appendix 3. Annex F. Report of the sub-committee on North Atlantic minke whales. Rep. Int. Whal. Comm. **41**: 132–171.

International Whaling Commission. 1992. Annex K. Report of the working group on North Atlantic minke trials. Rep. Int. Whal. Comm. **42**: 246–251.

International Whaling Commission 1994*a*. Report of the Scientific Committee. Annex H. The Revised Management Procedure (RMP) for baleen whales. Rep. Int. Whal. Comm. **44**: 145–167.

International Whaling Commission. 1994*b*. Report of the Scientific Committee. Annex J. Guidelines for conducting surveys and analysing data within the Revised Management Scheme. Rep. Int. Whal. Comm. **44**: 168–173.

International Whaling Commission. 2004. Report of the Scientific Committee. J. Cetacean Res. Manag. **6**(Suppl.): 107–154.

Jonsgård, Å. 1951. Studies on the little piked whale or minke whale (*Balaenoptera acutorostrata* Lacépède). Norsk Hvalfangsttidende, **40**: 80–95. [In Norwegian.]

Kuk, A.Y.C. 1995. Asymptotically unbiased estimation in generalized linear-models with random effects. J. R. Stat. Soc. Ser. B Methodological, **57**: 395–407.

Øien, N. 1988. Length distributions in catches from the Northeastern Atlantic stock of minke whales. Rep. Int. Whal. Comm. **38**: 289–295.

Øien, N. 1989. Sighting estimates of Northeast Atlantic minke whale abundance from the Norwegian shipboard survey in July 1987. Rep. Int. Whal. Comm. **39**: 417–421.

Okamura, H., Kitakado, T., Hiramatsu, K., and Mori, M. 2003. Abundance estimation of diving animals by the double-platform line transect method. Biometrics, **59**: 512–520.

Palka, D., and Hammond, P. 2001. Accounting for responsive movement in line transect estimates of abundance. Can. J. Fish. Aquat. Sci. **58**: 777–787.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. 1992. Numerical recipes in C. The Art of Scientific Computing. Cambridge University Press, Cambridge, UK.

Punt, A.E., Cooke, J.G., Borchers, D.L., and Strindberg, S. 1997. Estimating the extent of additional variance for southern hemisphere minke whales from the results of the IWC/IDCR cruises. Rep. Int. Whal. Comm. **47**: 431–434.

Ripley, B.D. 1977. Modelling spatial patterns (with discussion). J. R. Stat. Soc. Ser. B, **47**: 172–212.

Schweder, T. 1974. Transformation of point processes: application to animal sighting and catch problems, with special emphasis on whales. Ph.D. thesis, University of California, Berkeley, Calif.

Schweder, T. 1977. Point process models for line transect experiments. *In* Recent developments in statistics. *Edited by* J. Barra. North Holland Publishing Co., Amsterdam. pp. 221–242.

Schweder, T. 1997. Measurement error models for the Norwegian minke whale survey in 1995. Rep. Int. Whal. Comm. **47**: 485–488.

Schweder, T., and Hjort, N.L. 2002. Confidence and likelihood. Scand. J. Stat. **29**: 309–332.

Schweder, T., and Høst, G. 1992. Integrating experimental data and survey data to estimate $g(0)$: a first approach. Rep. Int. Whal. Comm. **42**: 575–582.

Schweder, T., Hagen, G., Helgeland, J., and Koppervik, I. 1996. Abundance estimation of Northeastern Atlantic minke whales. Rep. Int. Whal. Comm. **46**: 391–405.

Schweder, T., Skaug, H.J., Dimakos, X., Langaas, M., and Øien, N. 1997. Abundance estimates for Northeastern Atlantic minke whales. Estimates for 1989 and 1995. Rep. Int. Whal. Comm. **47**: 453–484.

Schweder, T., Skaug, H.J., Langaas, M., and Dimakos, X. 1999. Simulated likelihood methods for complex double-platform line transect surveys. Biometrics, **55**: 678–687.

Sigurjonsson, J., Gunnlaugsson, T., Ensor, P., Newcomer, M., and Vikingsson, G. 1989. North Atlantic sightings survey 1989 (NASS-89): shipboard surveys in Icelandic and adjacent waters July–August 1989. Rep. Int. Whal. Comm. **41**: 559–572.

Skaug, H.J., Øien, N., Bøthun, G., and Schweder, T. 2002. Abundance of Northeastern Atlantic minke whales for the survey period 1996–2001. Paper SC/54/RMP5, the Scientific Committee of the International Whaling Commission. Available from bemata.imr.no/whales.

Stockin, K., Fairbairns, R., Parsons, E., and Sims, D. 2001. Effects of diel and seasonal cycles on the dive duration of the minke whale (*Balaenoptera acutorostrata*). J. Mar. Biol. Assoc. U.K. **81**: 189–190.

Tierney, L., and Kadane, J.B. 1986. Accurate approximations for posterior moments and marginal distributions. J. Am. Stat. Assoc. **81**: 82–86.

Ugland, K. 1976. Population studies on eastern North Atlantic minke whales, *Balaenoptera acutorostrata* Lacepede. Rep. Int. Whal. Comm. (Sci. Rep.), **26**: 366–381.