

Research

Modeling Distribution and Abundance of Antarctic Baleen Whales Using Ships of Opportunity

*Rob Williams*¹, *Sharon L. Hedley*², and *Philip S. Hammond*²

ABSTRACT. Information on animal abundance and distribution is at the cornerstone of many wildlife and conservation strategies. However, these data can be difficult and costly to obtain for cetacean species. The expense of sufficient ship time to conduct design-unbiased line transect surveys may be simply out of reach for researchers in many countries, which nonetheless grapple with problems of conservation of endangered species, by-catch of small cetaceans in commercial fisheries, and progression toward ecosystem-based fisheries management. Recently developed spatial modeling techniques show promise for estimating wildlife abundance using non-randomized surveys, but have yet to receive much field-testing in areas where designed surveys have also been conducted. Effort and sightings data were collected along 9 650 km of transects aboard ships of opportunity in the Southern Ocean during the austral summers of 2000–2001 and 2001–2002. Generalized additive models with generalized cross-validation were used to express heterogeneity of cetacean sightings as functions of spatial covariates. Models were used to map predicted densities and to estimate abundance of humpback, minke, and fin whales in the Drake Passage and along the Antarctic Peninsula. All species' distribution maps showed strong density gradients, which were robust to jackknife resampling when each of 14 trips was removed sequentially with replacement. Looped animations of model predictions of whale density illustrate uncertainty in distribution estimates in a way that is informative to non-scientists. The best abundance estimate for humpback whales was 1 829 (95% CI: 978–3 422). Abundance of fin whales was 4 487 (95% CI: 1 326–15 179) and minke whales was 1,544 (95% CI: 1,221–1,953). These estimates agreed roughly with those reported from a designed survey conducted in the region during the previous austral summer. These estimates assumed that all animals on the trackline were detected, but preliminary results suggest that any negative bias due to violation of this assumption was likely small. Similarly, current methodological limitations prohibit inclusion of all known sources of uncertainty in the favored variance estimator. Meanwhile, our approach can be seen generally as an inexpensive pilot study to identify areas of predicted high density that could be targeted to: inform stratified designs for future line transect surveys, making them less expensive and more precise; increase efficiency of future photo-identification or biopsy studies; identify candidate time-area fisheries closures to minimize by-catch; or direct ecotourism activities. The techniques are likely to apply to areas where funding is limiting, where cetacean studies or wilderness-based tourism are just beginning, or in regions where even a very rough estimate of animal abundance is needed for conservation or management purposes.

Key Words: *abundance; Antarctic; baleen whale; cetacean; distance sampling; distribution; line transect; platform of opportunity; spatial model*

INTRODUCTION

Conservation of threatened species and management of exploited ones call for reliable information on distribution, abundance, and trends in abundance. Quantitative research on marine species is important for numerous, yet often conflicting, reasons. Marine mammals are often of management and conservation concern because they overlap spatially and ecologically with what humans want from the oceans (Trites et al. 1997). Cetaceans (whales, dolphins, and porpoises), with their vast ranges, represent a particular challenge for population monitoring (Talbot 1974, Donovan 1986, Bowen 1997, Berggren et al. 2002a). They spend small fractions of their time at the surface, and even then, only small portions of the animals' bodies are visible. Although arboreal marsupials are also cryptic and difficult to study (Lindenmayer et al. 2003), researchers can at least walk in their habitat. Pelagic cetacean research requires a ship, so these animals are not only difficult to study, but also costly to study (Burns and Wandesford-Smith 2002, Hammond et al. 2002).

Now consider, for example, the problem of by-catch of small cetaceans in commercial fishing nets. This problem is large, global, and growing (Northridge and Hofman 1999, Berggren et al. 2002b). Quantitative methods exist for assessing objectively whether observed levels of by-catch are biologically sustainable for a given population (Wade 1998), but these require information on population size and an estimate of the uncertainty associated with this parameter. Without these estimates, managers are in the unenviable position of having to balance the risk of ignoring a potential conservation threat on the one hand, against imposing unnecessary financial pressures (such as time-area closures, pingers or gear modification) on fishermen on the other. This may create a scenario where managers in affluent countries could choose to conduct surveys to improve the precision or accuracy of abundance estimates as by-catch levels deemed unsustainable are neared. However, such surveys may not be an option for researchers working in other areas, despite a similar need for information in order to manage or conserve marine resources (Vidal 1993).

Cetacean biologists commonly use line-transect sightings surveys in order to provide information on

abundance and distribution (e.g., Buckland et al. 2001). However, in their conventional form, line-transect surveys should be designed to give all points in the study area an equal probability of being sampled (Buckland et al. 2001), with transects being placed at random with respect to the distribution of the animals being studied. Conceptually, this is easy to do, but chartering a ship to conduct such a design-unbiased survey can be costly, sometimes prohibitively so.

When a survey design violates assumptions of equal coverage probability, other methods are required. One model-based approach is to use generalized additive models (GAMs, Hastie and Tibshirani 1990) to describe animal density along the trackline as smooth or linear functions of spatial or environmental covariates, and then use that relationship to predict density throughout the study area. Spatial modeling methods using line-transect survey data were developed by Hedley et al. (1999) and Bravington (2002). These techniques allow descriptive models of heterogeneity in whale density to be fitted to line-transect-type data, but without the requirement for random placement of tracklines in the study area. Data from such surveys could potentially be combined when fitting the models, thus imparting information on how animals use their habitat, and on how populations behave over time, even when sample sizes from individual surveys may be small.

An extreme case of surveys that fail to provide equal coverage probability is presented by those conducted from ships whose routes cannot be determined by the researcher, namely ships of opportunity. It would be useful to have inexpensive methods to allow coarse information on abundance and distribution to be obtained in cases where funds are not available for chartering a suitable ship to conduct a design-unbiased survey. Many researchers have attempted to glean information from so-called "Platforms of Opportunity," where the platform may be sighting logs (Braham and Dahlheim 1982, Moore et al. 1999), whaling records (Mizroch 1983, 1984, Gregr et al. 2000, Gregr and Trites 2001), or ships from which both effort and sightings were recorded (Northridge et al. 1995, Marques 2001). On opportunistic surveys where effort is recorded, not only is coverage non-random, but also it is generated by other people's needs, rather than a quantifiable algorithm (and thus one cannot simply weight observations based on sampling inclusion probability in a Horvitz-

Thompson-like estimator, e.g., Borchers et al. 1998a). Consequently, the concept of “coverage probability” in terms of survey design is not meaningful; the narrow covered region along the ship’s course has certain coverage probability, whereas everywhere else has zero coverage probability. Design-based methods, which rely on all points in the study area having equal or at least estimable probability of coverage, are therefore inappropriate for data from platforms of opportunity. These data can only be analyzed using a suitable model-based technique, such as the GAM-based spatial models presented in this paper.

Wilderness-based tourism is a global and growing industry (Nelson 1994). More than 10 000 tourists now visit Antarctica annually on expedition-style natural history cruises aboard small, ice-strengthened ships (International Association of Antarctic Tour Operators 2000). As a fleet, these ships provide reasonable coverage of the Scotia Sea and Antarctic Peninsula region, straddling parts of the International Whaling Commission (IWC) Areas I and II (Fig.1). Undertaking sightings surveys from a tourist ship that is already on the water is an attractive option for scientists wanting to study important areas that are otherwise prohibitively expensive to access. The tourism industry benefits as well from trained observers who can teach passengers about the area, and provide guests with the sense that they are facilitating research that will aid conservation of Southern Ocean wildlife. The importance of the Southern Ocean to global cetacean conservation cannot be overemphasized (Laws 1977). Commercial whaling removed approximately two million whales from the southern hemisphere between 1904 and 1980, representing 80–95% of all great whales in the region (Baker and Clapham 2002). Monitoring their recovery is clearly important.

We collected line-transect survey data on cetaceans from Antarctic ships of opportunity in the South Atlantic sector of the Southern Ocean during the 2000–2001 and 2001–2002 austral summers. Our primary goal was to provide an inexpensive description of abundance and distribution of three balaenopterid species: humpback (*Megaptera novaeangliae*), fin (*Balaenoptera physalus*) and Antarctic minke (*B. bonaerensis*) whales. Our secondary goal was to illustrate the uncertainty in our predicted animal distribution maps using custom animations.

METHODS

An expanded description of the methods used is found in App. 1 (including details about ship size and platform height in Table 1).

Data Collection

Data were collected on eight trips between 3 December 2000 and 14 March 2001, and on six trips between 5 December 2001 and 12 February 2002 (Table 2). All our trips started or ended in Ushuaia, Argentina (Fig. 1), and lasted from 6 to 24 days. Two ships were used during the first season, and a third ship in the second season.

Effort

Data were collected from the highest accessible point (the “primary platform”) on each of the three ships used in this study (Table 1). At the beginning and end of each data session, a handheld GPS unit was used for recording time, ship’s location, course, and speed. Information was recorded on factors that could affect sighting conditions, including sea state, cloud cover, and precipitation, and a subjective visibility code estimated the range at which a minke whale might be visible. Effort data were collected every 30 min thereafter, or more frequently if sighting conditions changed or if the ship made a marked change in course or speed. One of us (RW) was present during all data collection sessions. Consequently, although the ship represents a platform of opportunity, the effort and sightings were collected using standardized survey methods by a trained observer, rather than by tourists.

Sightings

When a cetacean school was spotted, it was assigned a sighting number. An angle board mounted on the deck railing was used to measure radial angle to the sighting, and a visual estimate was made of the range. Radial distance estimates were corrected subsequently using photogrammetric experiments described in Williams (2003). Ship location and the time of the sighting were recorded, and binoculars were used to confirm species and school size.

Fig. 1. Study area in the context of International Whaling Commission (IWC) Antarctic baleen whale areas. Most trips left from Ushuaia, Argentina or Port Stanley, Falkland Islands [Islas Malvinas]. Trips to the Antarctic Peninsula were more common than those to South Georgia, necessitating a post hoc designation of an Antarctic Peninsula–Drake Passage stratum (left of the dotted line) and a Scotia Sea (right of the dotted line) stratum.

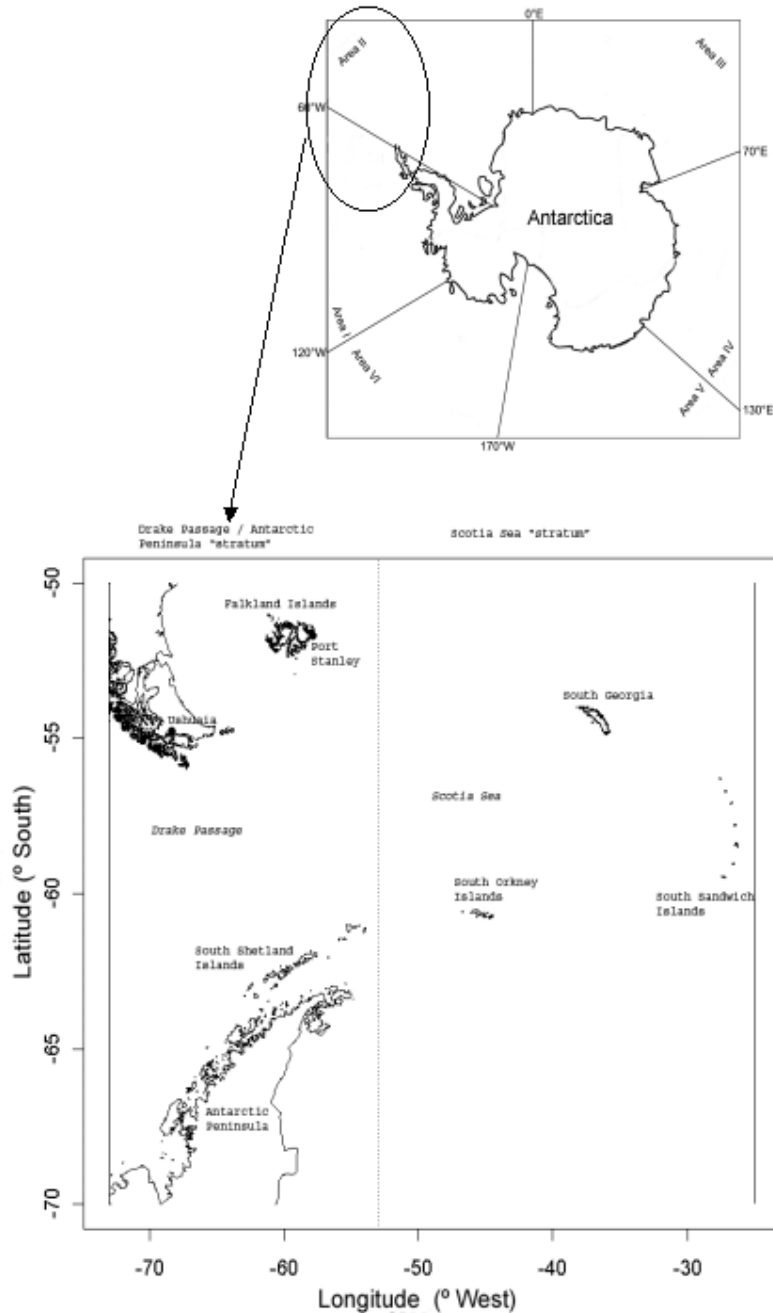


Table 1. Height of the primary observation and secondary (duplicate tracking) platforms aboard three ships of opportunity used in this study.

Ship	Platform height (m)	
	Primary	Secondary
<i>Mariya Yermolova</i>	14	12
<i>Lyubov Orlova</i>	14	12
<i>Explorer</i>	12	10

Data Analysis

Modeling detection probability and estimation of mean school size

Exploratory analysis of the sightings data was conducted in Distance 4.0 Release 2 (Thomas et al. 2002). Detection probability for each species was modeled as smooth functions of perpendicular distance (x). Candidate forms for the detection function were the hazard-rate and half-normal models (Buckland et al. 2001). Data were analyzed as though all whale schools directly on the trackline were spotted, that is, the so-called $g(0)=1$ assumption (Buckland et al. 2001). [Preliminary double-platform trials (Borchers et al. 1998b) showed that 92%–99% of these large animals were detected on the trackline, that is that $g(0)$ ranged from 0.92 for minke whales to 0.99 for fin whales, so violation of this assumption was unlikely to have introduced serious bias (Williams 2003).] School size and sea state were included as candidate covariates, and model selection was guided by Akaike’s Information Criterion (AIC; Akaike 1973). All other things being equal, large schools are detectable at greater distance than small schools. To address this, the default regression approach used in Distance 4.0 was used (Thomas et al. 2002). The selected detection function was extracted from Distance and used in R (<http://www.r-project.org>) for spatial modeling. These analysis methods are described in greater detail in App. 1.

Spatial Modeling from Line-transect Survey Data

Building descriptive models

The effort and sightings data were modeled using the “count” method (Hedley et al. 1999). Tracklines from each data collection session were divided into segments approximately 2 nautical miles (nm) in length. A previous simulation study (Hedley 2000) indicated that this was a suitable segment length to capture the heterogeneity in encounter rate for Antarctic baleen whales, but also that the count method was highly robust to variation in segment length. Start and end locations of the segments were calculated using the Geofunc add-in (Laake 2001) for EXCEL 2000®.

The location of the midpoint of each segment (latitude, LatMid, and longitude, LonMid) was calculated, along with two additional candidate explanatory variables. The first of these, depth, was calculated in ARCINFO® using a digital bathymetric database for the study area, DBDB-V Database <https://128.160.23.42/dbdbv/dbvquery.html>. The second variable was the distance from each segment to the nearest coastline, mindist. Spatial models do not model density directly largely because of problems linking density to an appropriate distributional form of the response variable. A solution to this problem involves including the area searched as an offset term in the model. School density in a segment, d_i , is estimated by:

Table 2. Search effort.

	Season	Ship	Start	Date	Primary	Secondary	Primary	Duplicate
1	2001	Yermolova	3 Dec	17 Dec	926	0	26	0
2	2001	<i>Yermolova</i>	20 Dec	27 Dec	411	0	11	0
3	2001	<i>Yermolova</i>	30 Dec	22 Jan	1298	462	87	40
4	2001	<i>Orlova</i>	24 Jan	30 Jan	180	0	1	0
5	2001	<i>Orlova</i>	2 Feb	9 Feb	342	0	11	0
6	2001	<i>Orlova</i>	11 Feb	16 Feb	297	0	2	0
7	2001	<i>Orlova</i>	19 Feb	26 Feb	680	0	27	0
8	2001	<i>Orlova</i>	1 March	14 March	1013	46	34	10
9	2002	<i>Explorer</i>	5 Dec	11 Dec	341	0	17	0
10	2002	<i>Explorer</i>	14 Dec	20 Dec	257	0	2	0
11	2002	<i>Explorer</i>	26 Dec	2 Jan	244	134	11	6
12	2002	<i>Explorer</i>	9 Jan	17 Jan	920	66	26	3
13	2002	<i>Explorer</i>	20 Jan	1 Feb	1745	171	80	9
14	2002	<i>Explorer</i>	4 Feb	12 Feb	1329	230	29	5
				Total:	9981	1109	364	73

$$\hat{d}_i = \frac{\hat{n}_i \cdot \hat{f}(0)}{2l_i} \quad (1)$$

where n_i is the number of detected schools in the segment, l_i is the length of the segment, and

$$\hat{f}(0) \quad (2)$$

is the estimated probability density function evaluated at zero perpendicular distance (Burnham and Anderson 1976). By including the estimated covered area of each segment as an offset term in the model, *offset* =

$$\log \left(\frac{2l_i}{\hat{f}(0)} \right) \quad (3)$$

the response variable,

$$\hat{N}_s \quad (4)$$

can be modeled as count data. It would be unrealistic to expect whale schools to be spread randomly throughout the region, so these counts should not be modeled using a Poisson error distribution (where the variance of each observation is assumed to be equal to its mean). As in other similar studies (e.g., Clarke et al. 2000, Forney 2000, Hedley 2000, Marques 2001), we model the clustering in the data

using quasi-likelihood, assuming that variance is proportional to the mean, and thus accounting for at least some of the over-dispersion present in the data (compared with a Poisson distribution). Although this technique does not alter the point estimates compared with a Poisson assumption, it should provide more realistic variance estimates. Nevertheless, some residual small-scale, localized correlation is likely to remain unmodeled. The model was thus of the following general form:

$$E(n_i) = \exp \left[\ln \left(\frac{2l_i}{\hat{f}_0} \right) + \beta_0 + \sum_k s_k(z_{ik}) \right] \quad (5)$$

where:

$$\beta_0 \quad (6)$$

is a parameter to be estimated; and z_{ik} represents the value of the k^{th} explanatory spatial variable in the i^{th} segment, which is a smooth function, s_k , of the explanatory variable.

Note that if detectability varies within a segment (e. g., because of varying sea states) this model may be further generalized to:

$$E(n_i) = \exp \left[\ln(2l_i w_i \hat{p}_i) + \beta_0 + \sum_k f_k(z_{ik}) \right] \quad (7)$$

where:

w_i is the distances at which perpendicular sightings were truncated in the i^{th} segment; and

$$\hat{p}_i \quad (8)$$

is the estimated probability of detection of a school in the i^{th} segment (i.e., estimated from f , which could vary between segments according to the value of covariates).

The models were fitted in the package *mgcv* (Wood 2001) available within the R software (R Development Core Team 2003). Flexibility and model term dropping were guided by the framework recommended by Wood (2001, App. 1). Note that *mgcv* uses Generalized Cross Validation for model selection. Thus, the procedure automates a training and cross-validation approach for choosing a statistically defensible degree of smoothing, with penalties for unnecessary flexibility.

Defining the Study Area

The study area is difficult to define a priori for surveys conducted from ships of opportunity. In this study, the tracklines themselves were used for post hoc definition of the study area. A convex hull was fitted around the tracklines in the Antarctic Peninsula–Drake Passage stratum (Strindberg 2001; Fig. 1), where the bulk of the search effort occurred and within which area coverage appeared reasonable. This defined the area across which whale density would be predicted. Although sightings and effort data from the Scotia Sea stratum (Fig. 1) were included in all descriptive modeling steps, no attempt was made to predict in that stratum, because coverage probability in the middle of that stratum was zero. Including effort beyond the study area provides good structure to the descriptive model, and tames the tendency for flexible models to extrapolate unrealistically high density at the peripheries of study areas where coverage is poor—a so-called “edge effect” (Clarke et al. 2000, Bravington 2002). However, including this region in the prediction grid’s convex hull would have approached extrapolation rather than interpolation, because all trips in our study were restricted to the periphery of the Scotia Sea.

Predicting Whale Density throughout the Study Area

A gridded data set was created, containing a value in every grid cell for each explanatory variable in the model. A square grid size of 5 km on a side (i. e., 25 km²) was chosen for prediction. This was

arbitrary, but constrained by two requirements: the resolution had to be coarser than the segment length (2 nm), but not large enough for values of the explanatory variables to vary much within the cell. Values for the explanatory variables (latitude, longitude, depth, and distance offshore) were calculated using the value at the midpoint of each grid square. Extensive simulations (Hedley 2000) revealed that model descriptions and predictions were robust to variation in choice of grid size, but the area of the grid square should be larger than that of segments along the trackline.

The prediction grid data were passed to the descriptive model selected for each species using the `predict.gam` function in `mgcv`. The output of the model was an estimate of the predicted number of whale schools in each grid cell, based on each cell's latitude, longitude, distance offshore, depth, and area. This predicted count was converted to density for mapping purposes by dividing the count by the area of each cell. Animal abundance was calculated by multiplying the predicted density in each cell by expected school size (from the size-bias regression in the detection function modeling step; Buckland et al. 2001) and by the area of each cell, and taking the sum of all values in the grid.

Estimating Variance of Abundance Estimates

In conventional distance sampling surveys, variance may be estimated from the sample variance of independent transect lines (Buckland et al. 2001). However, not only is it difficult to identify replicate transects from data collected aboard ships of opportunity (Marques 2001), but also the segments of effort used in the spatial model cannot be assumed to be independent, so analytical variance estimates are inappropriate. Buckland et al. (2001) recommended resampling the effort and sightings data to obtain variance estimates in such cases.

Jackknife estimates of variance have been shown to perform well in some spatial modeling applications (Hedley 2000), but not universally so. The jackknife may be thought of as an approximation of the nonparametric bootstrap (Efron and Tibshirani 1986), but is likely to be biased unless the number of resampling units is large. On the face of it, there may be a large number of sampling units in our study (e.g., days of search effort), but such units are not likely to be independent. Although the number of trips is small (14), each individual trip can be

thought of as independent, and in the absence of a better alternative, the method we used to estimate variance was the jackknife (Miller 1974, Efron 1979, Efron and Stein 1981), using each trip as the resampling unit. Other resampling-type estimators of variance were found to perform poorly (Williams 2003). Therefore, variance was estimated by removing each trip's effort and sightings data in turn, with replacement, and analyzing the remaining data to predict abundance. The 95% confidence intervals of these 14 pseudovalues were calculated using standard jackknife estimator methods assuming a log-normal distribution (Buckland et al. 2001).

RESULTS

Searching Effort

Searching effort totaled approximately 10 000 km (Table 2). Ship speed averaged 12 kn (23 km h⁻¹, 0.15SE) during the survey. Searching was carried out in Beaufort 4 conditions or better along 78% of the trackline. Visibility was excellent (>2 km) for 94% of the search effort.

Sightings

A total of 364 sightings of cetacean schools was observed. These are summarized by species in Table 3. The minimum number of sightings recommended for modeling detection probability is 60–80 schools (Buckland et al. 2001). Only three cetacean species were seen this frequently: humpback—129 schools, 232 individuals; fin—80 schools, 207 individuals; and minke—75 schools, 119 individuals. Consequently, subsequent analyses were restricted to these three species. An additional five schools (approximately nine animals) were spotted, but not identified to species. These were not included in the analyses.

Humpback Whales

Exploratory data analysis suggested truncating the perpendicular sighting distances at 2500 m. This truncation distance was used to assess the fit of different candidate detection functions. The model that fitted the resulting humpback sightings data best, as determined by AIC, was a hazard-rate key

Table 3. Sightings of cetacean schools by species.

Species	Schools	Indiv.
Humpback whale (<i>Megaptera novaeangliae</i>)	129	232
Fin whale (<i>Balaenoptera physalus</i>)	80	207
Minke whale (<i>B. bonaerensis</i>)	75	119
Hourglass dolphin (<i>Lagenorhynchus cruciger</i>)	19	86
Killer whale (<i>Orcinus orca</i>)	14	61
Southern bottlenose whale (<i>Hyperoodon planifrons</i>)	14	34
Peale's dolphin (<i>L. australis</i>)	9	46
Dusky dolphin (<i>L. obscurus</i>)	4	15
Southern right whale (<i>Eubalaena australis</i>)	4	7
Sei whale (<i>B. borealis</i>)	4	11
Long-finned pilot whale (<i>Globicephala melas</i>)	2	38
Cuvier's beaked whale (<i>Ziphius cavirostris</i>)	2	4
Sperm whale (<i>Physeter macrocephalus</i>)	2	6
Strap-toothed whale (<i>Mesoplodon layardii</i>)	1	1
Unidentified cetacean spp.	5	9

function with no adjustment terms. There was insufficient evidence in the data (based on AIC) to justify including sea state as a covariate. Effective strip half-width (esw) was estimated to be 1142 m (SE = 105). Mean group size was estimated by a size-biased regression to be 1.83 (SE = 0.07).

Modeling density along the trackline

The selected model included the following explanatory terms:

$$s(\text{LatMid}, 1.47) + s(\text{mindist}, 4) + s(\text{depth}, 4) + s(\text{LonMid}, 3.87)$$

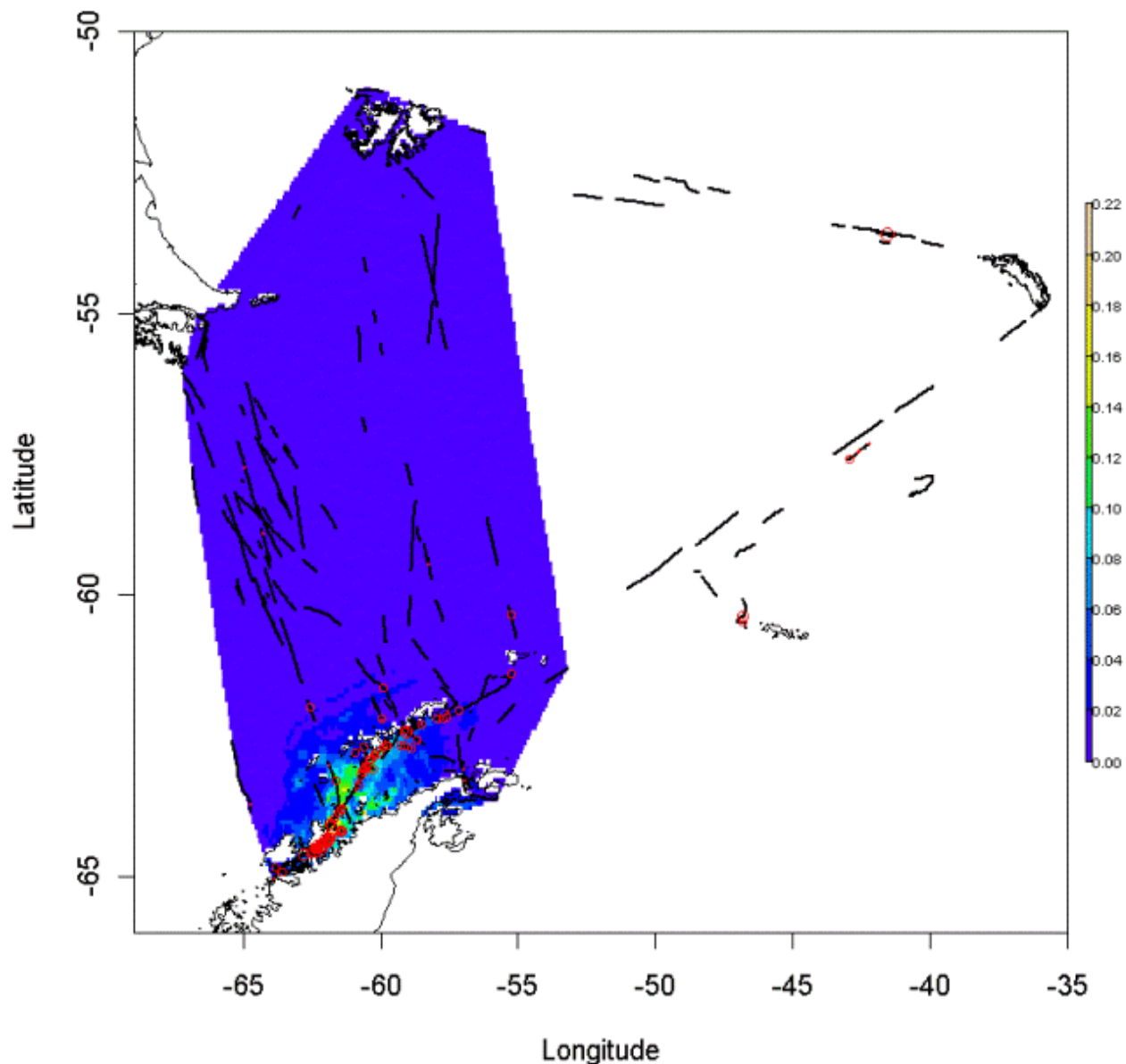
where the numbers in parentheses refer to the estimated degrees of freedom for each term selected by mgcv.

The effect of each variable on density, conditional on the other variables being included in the model, is shown in App. 2. The explanatory power of the model was moderate; the adjusted *R*-square score (Wood 2001) for the model was 0.129, and deviance explained was 36.1%.

Modeling humpback whale density and abundance across the prediction grid

The model predicted a very strong density gradient with the highest-density region predicted along the west side of the Antarctic Peninsula (Fig. 2). Humpback whale abundance in the best-covered region of the study area was estimated to be 1829 animals.

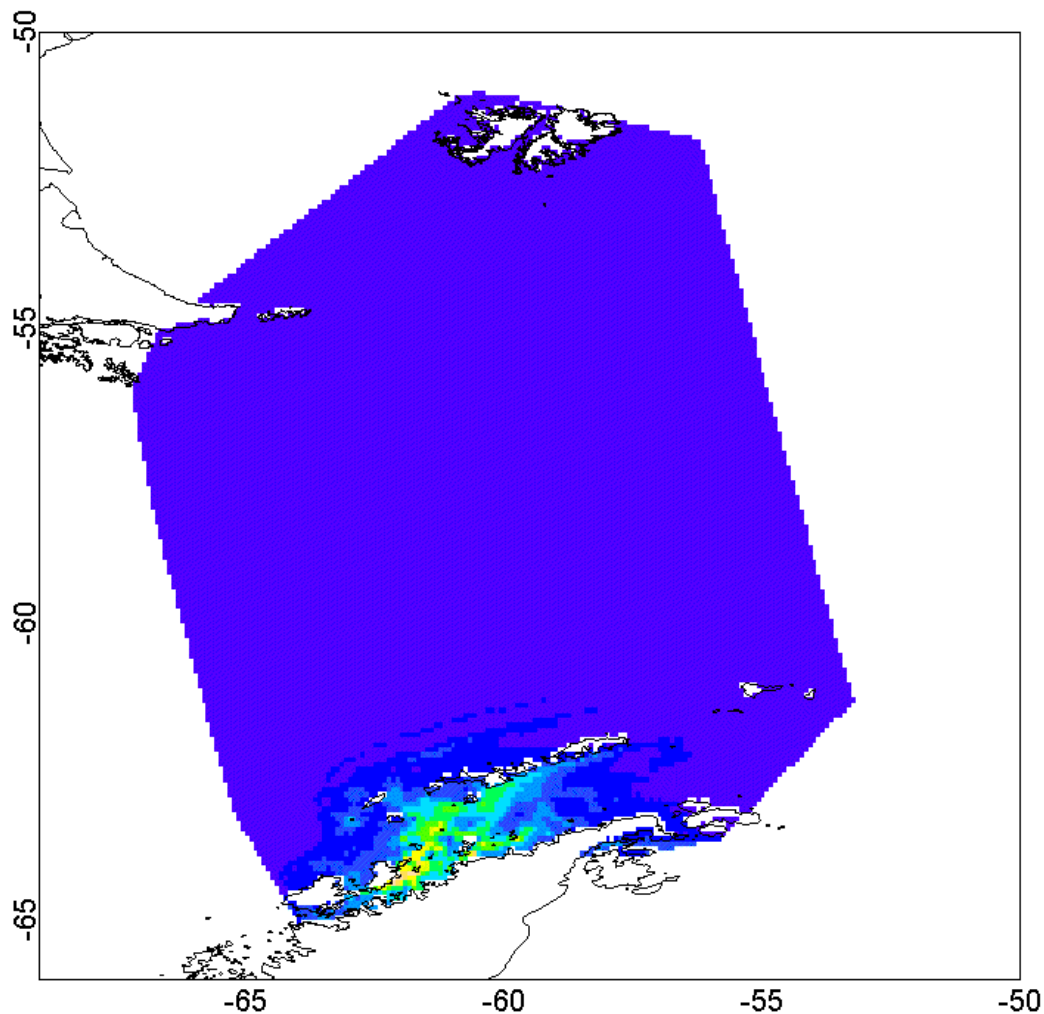
Fig. 2. Predicted density gradient of humpback whale schools in the best-covered region of the study area. The scale bar on the right is in units of schools per nm^2 . Tracklines are shown as solid black lines, and humpback sightings shown as red circles with radius proportional to school size.



The log-normal 95% confidence interval on humpback abundance was 978–3422. Predicted density gradient maps of humpback whale schools were created at each jackknife iteration, and are shown in Fig. 3. The area predicted to have the

highest density of humpback whale schools, the western Antarctic Peninsula region, was identified consistently at each iteration.

Fig. 3. An animation showing variation in predicted density of humpback whales. Each frame of the animation reflects an iteration of jackknife resampling: the removal of one trip's effort and sightings data; refitting the detection function on the remaining sightings; reevaluating size-bias corrected estimates of mean school size; refitting the spatial model; and prediction across the grid. White patches within the regions of high density reflect grid squares predicted to have >0.22 humpback whale schools/nm². The longest axis of the prediction grid (from the southern tip of South America to the northern tip of the Antarctic Peninsula) is approximately 500 nm long. Note that, although the scale of the density gradient varies among iterations, the placement of the highest-density area along the western Antarctic Peninsula is consistent.



Fin Whales

Exploratory data analysis suggested truncating the perpendicular sighting distances at 2000 m. This truncation distance was used to model the detection function, and the model that fitted the fin whale sightings data best, as determined by AIC, was a hazard-rate key function with no adjustment terms and without sea state as a covariate. Effective strip half-width was estimated to be 816 m (SE = 108). The regression-based estimate of mean school size was 2.44 (SE = 0.20).

Modeling density along the trackline

The selected model included the following explanatory terms:

$s(\text{LonMid}, \text{LatMid}, 1.47) + s(\text{mindist}, 3.93) + s(\text{depth}, 3.80)$

where the first term represents a two-dimensional smooth of latitude and longitude. The effect of each variable on density, conditional on the other variables being included in the model, is shown in App. 3. Fin whale density tended to be low in water depths shallower than 1000 m, and within 40 nm from the nearest coastline. The explanatory power of the model was moderate; the adjusted R -square score for the model was 0.103, and deviance explained was 37.9%.

Modeling density and abundance across the prediction grid

The model predicted a strong density gradient with the highest-density region predicted in the Scotia Sea (Fig. 4). Fin whale abundance in the best-covered region of the study area was estimated to be 4487 animals.

The log-normal 95% confidence interval on fin whale abundance was 1326–15 179. Predicted density gradient maps of fin whale schools were created at each jackknife iteration, and are shown in Fig. 5. Note that the predicted density of fin whales was substantially more variable than that for humpback whales, likely reflecting an edge effect related to poor coverage in the highest-density region of the study area.

Minke Whales

Exploratory data analysis revealed no long tail to the perpendicular distance distribution, so no truncation of perpendicular sighting distances was needed in order to obtain an adequate fit. The detection function that fitted the minke whale sightings data best, as determined by AIC, was a hazard-rate key function with no adjustment terms. Effective strip half-width was estimated to be 663 m (SE = 91). The regression-based estimate of mean school size was 1.209 (SE = 0.064).

Modeling density along the trackline

The selected model included the following explanatory terms:

$\text{LatMid} + s(\text{mindist}, 3.03) + s(\text{depth}, 2.01) + s(\text{LonMid}, 3.42)$

i.e., latitude entered the model as a linear term (which has one degree of freedom). The effect of each variable on density, conditional on the other variables being included in the model, is shown in App. 4. Minke whale density showed bimodal longitudinal peaks near 45°W and 65°W. The linear term of LatMid had a slope of $-0.30 (\pm 0.12\text{SE}, p = 0.011)$, indicating that the marginal effect of latitude on density was linear in a southward direction. The explanatory power of the model was quite poor, with an adjusted R -square score for the model of 0.105, and deviance explained of 23.4%. In particular, the model described the generally higher density regions around the Peninsula quite well, but provided an inadequate fit in the Drake Passage, where minke whale sightings were made, but no patches of higher density were predicted to occur. Because of the high number of segments with no sightings in the Drake Passage, the models were not sufficiently flexible to fit the local concentrations seen there well; rather the aim was to obtain maps of the general trend in density in the region as a whole. An increase in the flexibility allowed for each smooth term in the model may have produced a better fit, but may also have led to unrealistically high variance estimates.

Modeling density and abundance across the prediction grid

The model predicted a strong density gradient with a predicted high-density region spread along the Antarctic Peninsula (Fig. 6). Minke whale

Fig. 4. Predicted density gradient of fin whale schools in the best-covered region of the study area. The scale bar on the right is in units of schools per nm^2 . Tracklines are shown as solid black lines, and fin whale sightings are shown as red circles with radius proportional to school size.

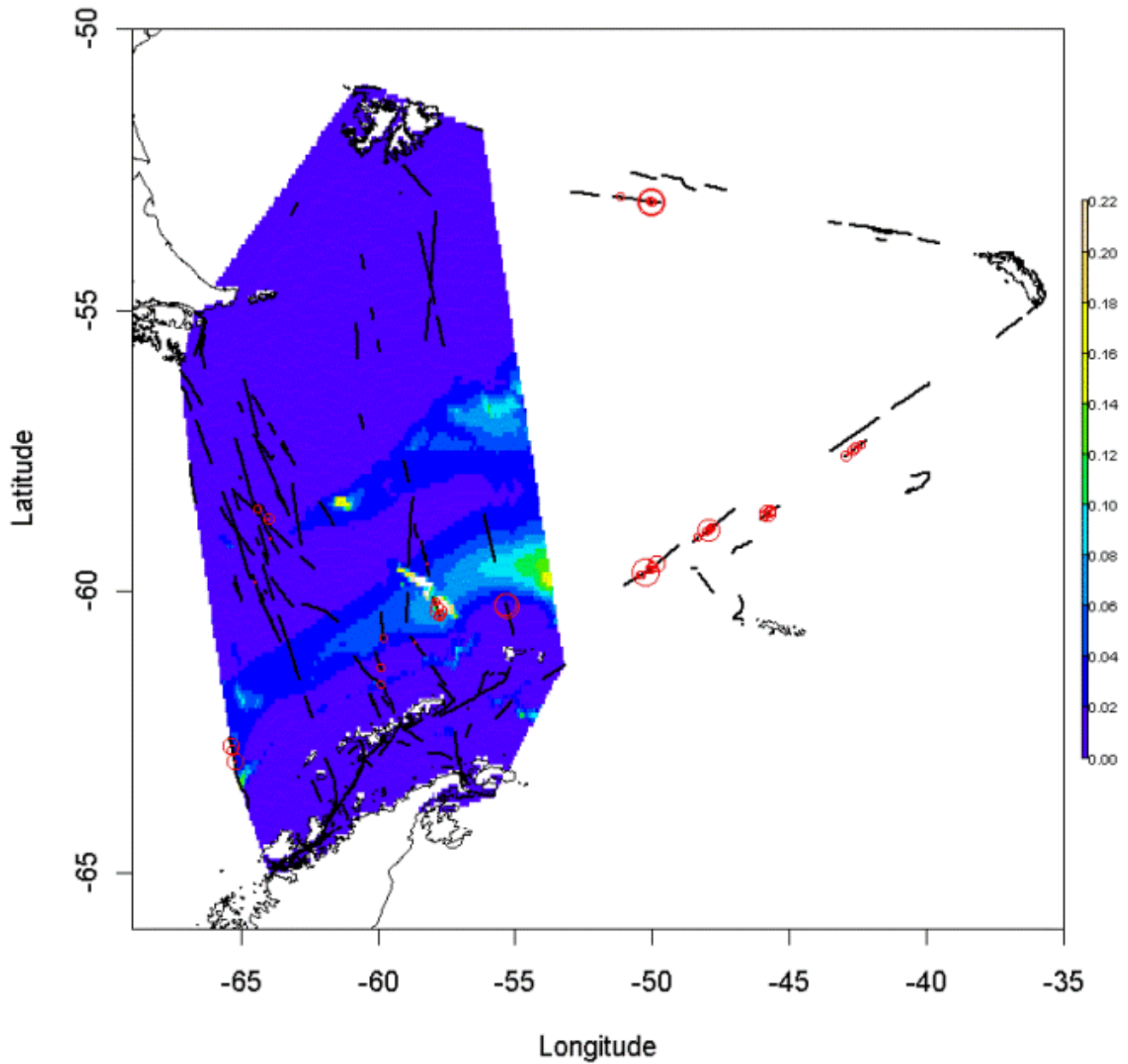


Fig. 5. An animation showing variation in predicted density of fin whales at each iteration. White patches within the regions of high density reflect grid squares predicted to have >0.22 fin whale schools/nm². The longest axis of the prediction grid (from the southern tip of South America to the northern tip of the Antarctic Peninsula) is approximately 500 nm long. Note that the area predicted to have the highest density of fin whale schools showed more variability than was seen in humpback whales (Fig. 2), but was always predicted to lie in offshore waters, running parallel to the Antarctic Peninsula.

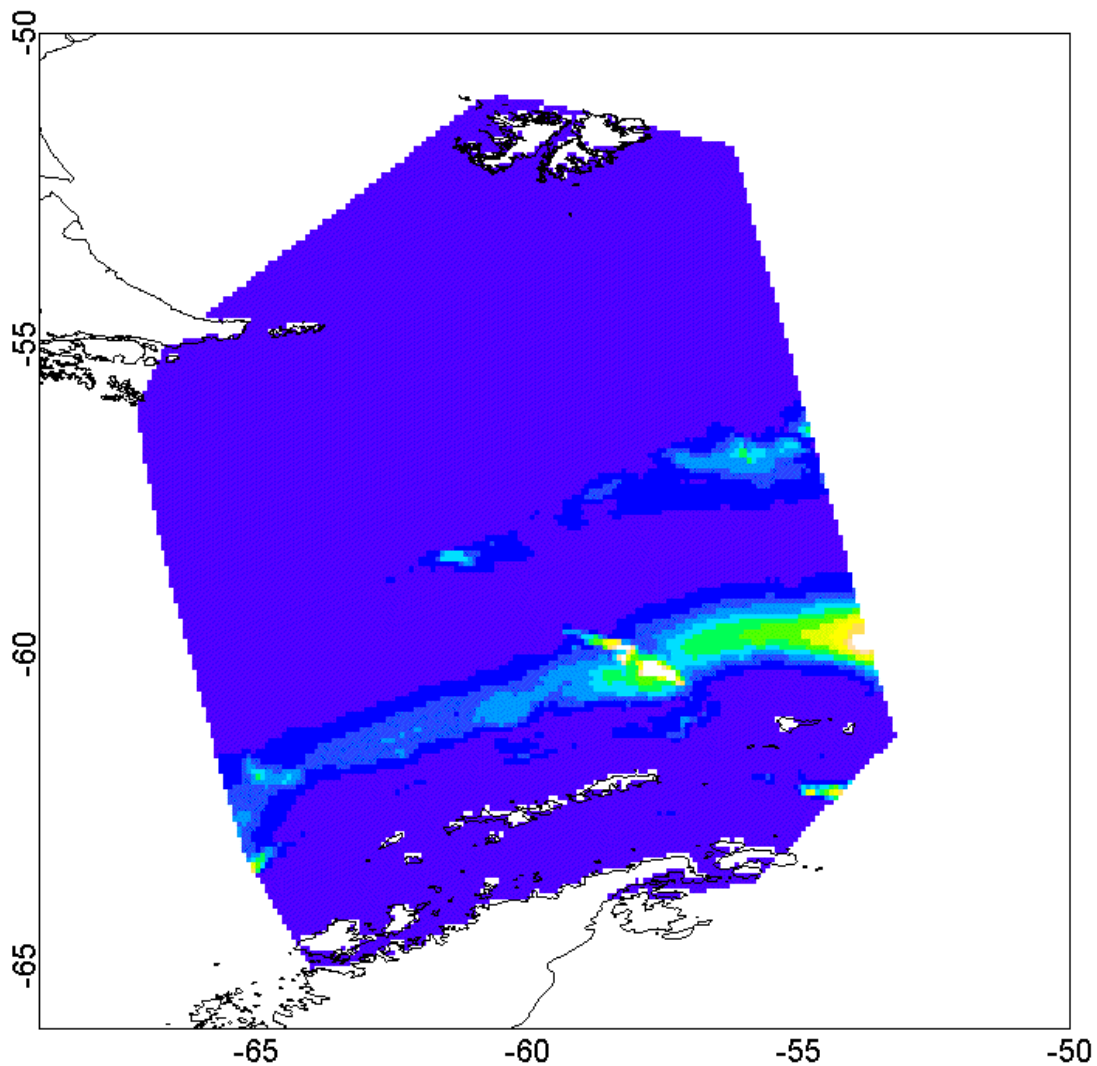


Fig. 6. Predicted density gradient of minke whale schools in the best-covered region of the study area. The scale bar on the right is in units of schools per nm^2 . Tracklines are shown as solid black lines, and minke sightings are shown as red circles with radius proportional to school size.

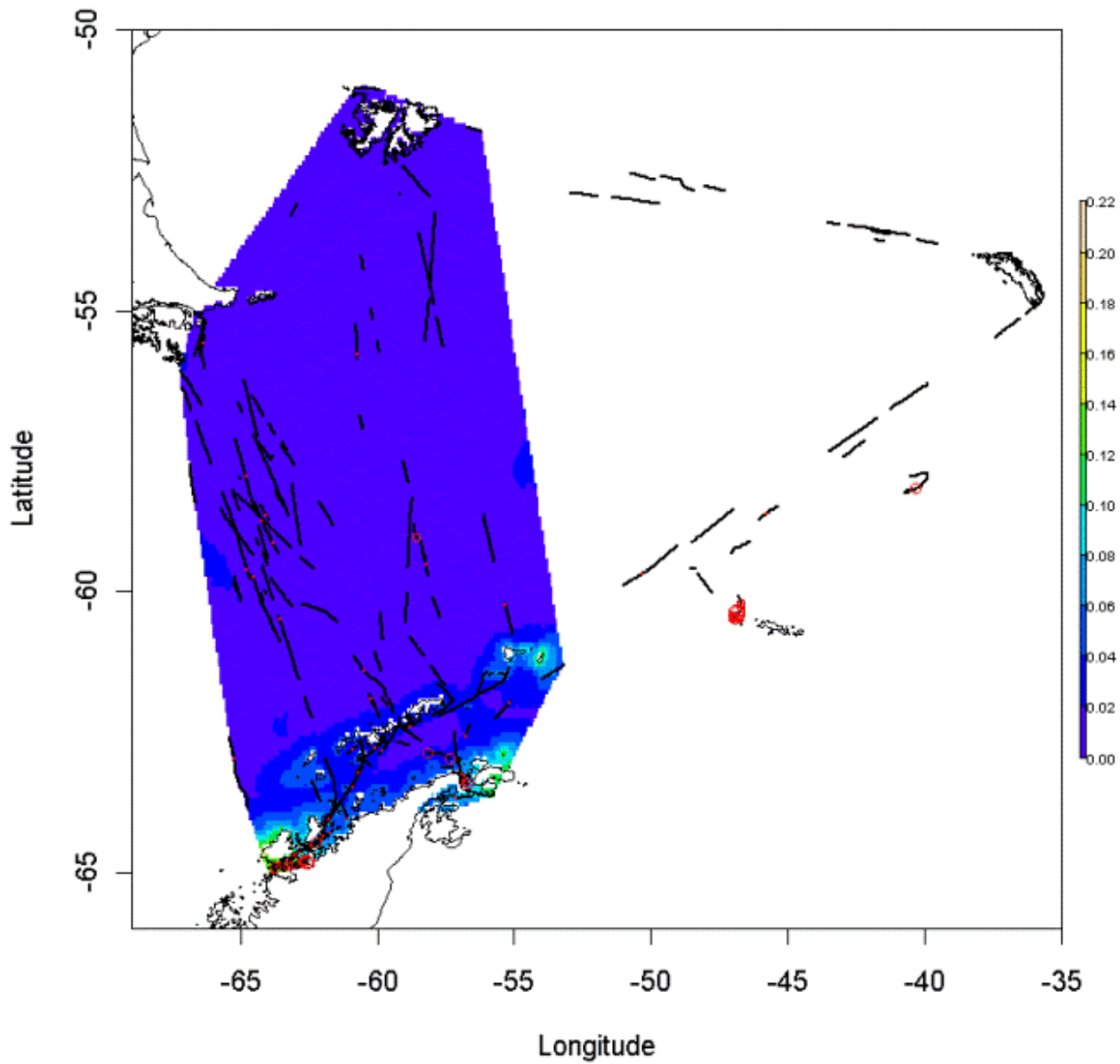
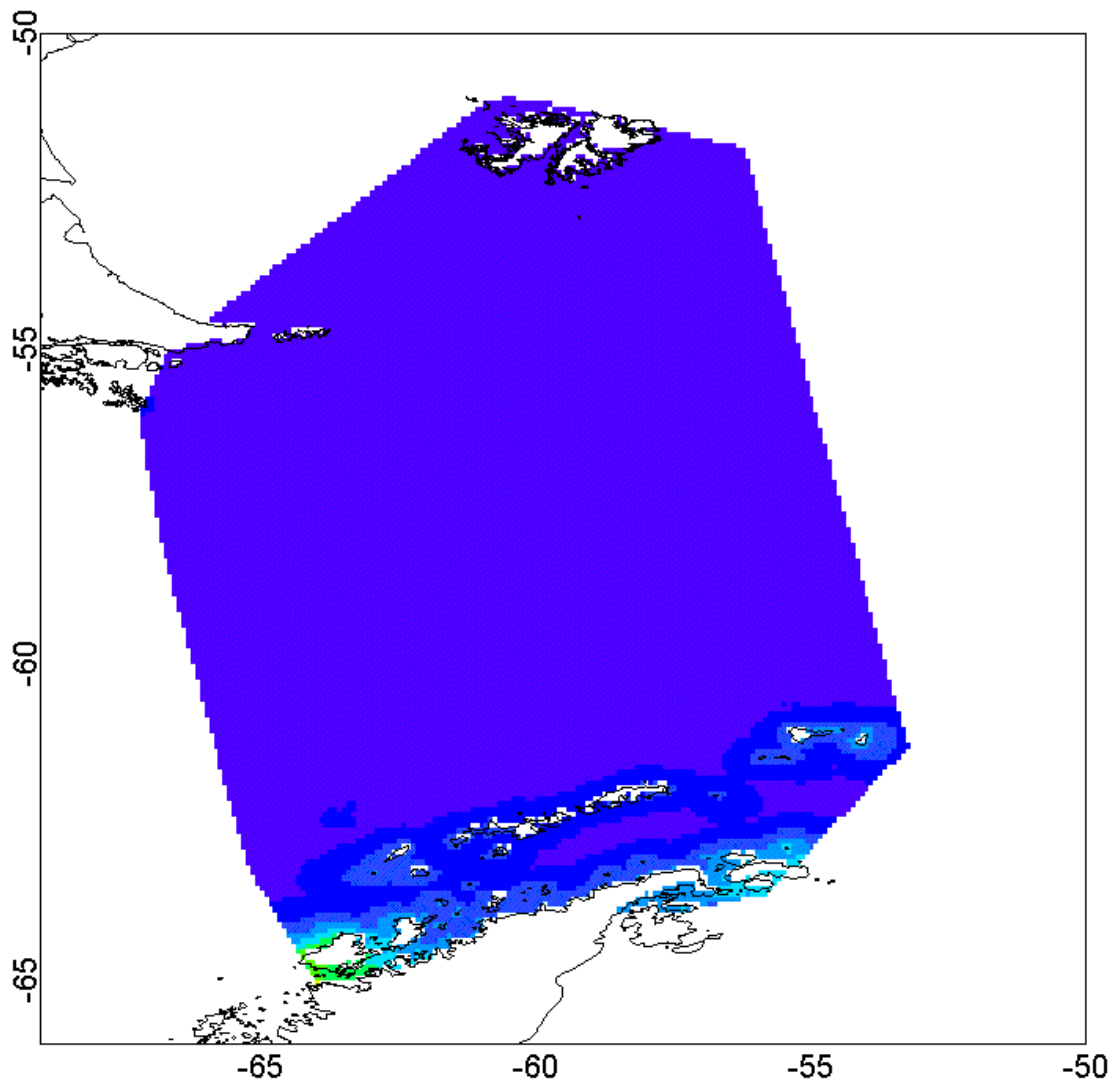


Fig. 7. An animation showing variation in predicted density of minke whales at each iteration. All grid squares were predicted to have <0.22 minke whale schools/nm². The longest axis of the prediction grid (from the southern tip of South America to the northern tip of the Antarctic Peninsula) is approximately 500 nm long. Areas close to shore, and in the icy waters off the northern tip of the Antarctic Peninsula were predicted to have high densities of minke whales in each iteration.



abundance in the best-covered region of the study area was estimated to be 1544 animals.

The log-normal 95% confidence interval on minke whale abundance was 1221–1953. Predicted density gradient maps of minke whale schools were created at each jackknife iteration, and are shown in Fig. 7. These predictions consistently identified high-density areas very close to shore, and off the northern tip of the Antarctic Peninsula (Antarctic Sound).

DISCUSSION

Spatial modeling of line-transect survey data collected from ships of opportunity yielded abundance estimates and predicted density maps for three baleen whale species in the Antarctic. Our looped animations of predicted animal density represent a novel way to illustrate uncertainty in estimates of both abundance and distribution to interested stakeholders who are non-scientists. This study demonstrates that there is merit in collecting reliable distance sampling data from a non-randomized survey with reasonable coverage, modeling heterogeneity along the trackline, and using the model to predict density throughout the study area. The framework outlined here is an appropriate way to gain useful information on frequently seen cetacean species in other areas from expedition-style cruise ships, fishing boats, freighters, or other ships of opportunity in understudied areas, or where lack of research funding prevents researchers from conducting design-unbiased surveys.

Distribution of Antarctic Balaenopterids and Other Cetaceans

Maps of cetacean distribution can be useful for guiding future research. The smooth density surfaces predicted for humpback, fin, and minke whales (Figs. 2, 4, and 6, respectively) could be used in a variety of ways, providing that they accurately reflect mean animal distribution during the survey. Sequential exclusion of each of the 14 trips affected the scale of the density gradient, but not its general orientation (Figs. 3, 5, and 7). Information on whale distribution and density plays a key role in management and conservation. For example, if local competition were suspected between commercial krill harvest and krill-eating predators (Laws 1977),

krill catch quotas could be allocated spatially in order to minimize it. One framework for assessing evidence for competition was outlined by Boyd (2002).

More generally, the identification of high-density areas could assist efforts to minimize by-catch of threatened species, or to set aside a protected area that satisfies the energy requirements of the population (Hooker et al. 1999). Places where whales are reliably seen can form core areas for wilderness-based tourism. Whalewatching is an important and growing industry worldwide (Hoyt 1997). Well managed, it promises to be a sustainable source of income for many displaced fishermen and for members of remote coastal communities.

But identification of high-density areas also plays an important iterative role in cetacean research itself. Studies that do not require a randomized sampling design routinely benefit from identifying high-density areas that can be targeted in future. This would increase the efficiency of photo-identification and biopsy studies. Identifying core areas, and directing research effort within them, were important contributions to the well-known and long-term studies of resident killer whale (*Orcinus orca*) populations of the northeast Pacific (Bigg et al. 1990, Baird and Whitehead 2000, Ford et al. 2000). Similarly, studies on spatial and temporal distribution (Hooker et al. 1999) and association patterns (Gowans et al. 2001) of northern bottlenose whales (*Hyperoodon ampullatus*) benefited from finding and targeting a high-density area, The Gully, off Nova Scotia, Canada. Spatial modeling can be thought of as a rapid and objective way to identify such hotspots.

Finally, pilot studies such as this can be used to map density gradients that aid the design of future line-transect surveys. By surveying across known density gradients, one avoids conflating real spatial variation in animal density with sampling variability. Knowledge of animal distribution can also guide stratification of future line-transect survey designs, with more effort being allocated to high-density areas. Informed decisions about transect placement and stratification will generally increase the precision of abundance estimates from future surveys. As a result, distribution maps such as the ones we present can be an end point in themselves or simply an aid to future survey design, depending on the research questions being asked, and on the financial resources at hand.

Abundance of Antarctic Balaenopterids

It is important to know, but difficult to assess, whether the absolute abundance estimates predicted by our spatial models were unbiased. Irrespective of survey design, the most obvious form of negative bias in any line-transect survey is the assumption that individuals at zero perpendicular distance were detected with certainty. Violation of this so-called $g(0)=1$ assumption can occur in two ways: not all animals were available to be detected (“availability bias”); or observers failed to detect all available animals (“perception bias,” Marsh and Sinclair 1989). Consequently, the $g(0)$ assumption is likely to hold strictly true only for highly visible and frequently surfacing animals, whose exhalations make them detectable at the surface several times while within an observer’s field of view. We referred earlier to our experimental data (Williams 2003), which indicated that detection probability for animals surfacing on the trackline actually ranged from 0.92 (minke whales) to 0.99 (fin whales). Consequently, perception bias was likely to have resulted in only minor underestimation of absolute abundance.

Similarly, our abundance estimates are likely to be biased low because availability bias is not accounted for, but we believe that this bias is likely to be small as well. Diving data can be used to test this contention. A recent review of common (i.e., *Balaenoptera acutorostrata*, rather than Antarctic, *B. bonaerensis*) minke whale diving behavior by the IWC Scientific Committee reported a mean surfacing rate of 49.2 s-h (IWC 2005 (App. 6)). In our study, minke whales were detected at a maximum radial distance of 2950 m, and a mean of 760 m. At an average ship speed of 12.6 kn, this corresponds to minke whales surfacing a maximum of seven times and an average of two times within the field of view. Thus, availability bias was likely to be low. The double-platform experiments discussed previously (Williams 2003) revealed that observers detected 92% of minke whales that were visible on the trackline. Thus, minke whales presented observers with several opportunities to detect them (i.e., availability bias was small), and when whales were at the surface, observers rarely missed detecting them (i.e., perception bias was small).

Although humpbacks (Dolphin 1987) and fin whales (Jahoda et al. 2003) surface less frequently on average than minke whales do, their blows are visible at a much greater distance. Using the steps

outlined above, we calculated roughly that observers had one to four opportunities to detect humpbacks and fin whales within the average and maximum field of view, respectively. Our double-platform data suggest that the probability of detecting a whale that surfaced on the trackline was even higher for fin and humpback whales (0.95 and 0.99) than for minke whales (0.92). Thus, availability and perception bias were likely very small for these species as well.

These very rough estimates were calculated using respiration data from whales in the northern hemisphere, but the exercise suffices to demonstrate our point, that negative bias was likely to be small for these species. The same would not hold true for less conspicuous species, or for animals that dive much longer than these three balaenopterid species. Indeed, $g(0)$ will be lower for more cryptic or longer-diving species, or under poorer survey conditions. For example, Barlow and Sexton (1996) reported an estimate of $g(0) = 0.87$ for the deep-diving sperm whale (*Physeter macrocephalus*). As these emerging techniques continue to be developed, however, statisticians will incorporate robust ways of addressing missed trackline detection into spatial modeling methods (Bravington 2002, 2004).

Until such new methods are available and proven, we used two subjective approaches to gauge the accuracy of our point estimates of abundance. The first was a series of internal checks. Point estimates of abundance varied by only 10%–20% when adding and dropping model terms, considering two-dimensional vs. one-dimensional smooths, or increasing model flexibility. The second, external approach to gauging accuracy of these abundance estimates was to compare estimates with those from designed surveys. There is no “ground truth” against which these estimates could be judged, because the study area was defined post hoc from the search effort (and, therefore, was not designed explicitly to be comparable with other surveys), and because true abundance is rarely known.

But one suitable data set for comparison is that collected on the CCAMLR-IWC Krill Synoptic Survey in January–February 2000, the year before the first season in our study (Hedley et al. 2001). The CCAMLR-IWC cetacean data were collected along a grid of systematically placed tracklines across a large section of the South Atlantic sector of the Southern Ocean. The CCAMLR-IWC study

covered a roughly rectangular region from approximately 49°W to 71°W, and from 60°S to 67° S. Thus, the earlier study area was roughly 2.5 times the size of ours, but located in the southern half of our study area (Fig. 1). The region of overlap between the two studies was found to be an area of high density for minke and humpbacks, and a region of relatively low density for fin whales. Given this, we would expect that the CCAMLR study would yield abundance estimates a few times greater than ours for humpback and minke whales. We would expect our fin whale estimate to be higher than the CCAMLR estimate, because our study sampled north of 60°S, and found areas of high fin whale density.

A spatial model was developed to model humpback, fin, and minke whale distribution from the CCAMLR-IWC data (Hedley et al. 2001). In fact, the CCAMLR-IWC estimate of 6991 animals was approximately 3.8 times our estimate of 1829. The CCAMLR-IWC estimate of 7395 minke whales was 4.8 times as high as our estimate of 1544. Our fin whale estimate of 4427 whales was approximately three times as high as the CCAMLR-IWC estimate of 1492. These calculations represent very simplistic comparisons of very complex systems: indeed, not all whales return to the feeding grounds annually, so neither study's estimates should be interpreted as "truth." However, in all three cases, the differences were in the direction of the trend predicted. And the differences are on the order of magnitude that one would expect given estimated differences in study area size, and observed interspecific differences in animal density. Thus, we conclude that our methods provided a rough but unbiased estimate of the average whale abundance in the region during the two austral summers of the study. Including both years in the study was necessary because of sample size constraints, but may have introduced additional variability. Ongoing work using these platforms will allow us to assess in future inter-annual variability in whale density in this region, as well as factors that influence inter-annual variability in animal distribution.

It should be noted that this study aimed to provide estimates of absolute abundance. An alternative use of these platforms might have been to collect estimates of relative abundance (e.g., presence/absence, encounter rate, or density). The use of platforms of opportunity for collecting information on relative abundance is well established, and it has

been shown that opportunistic platforms (or other indirect measures of abundance such as levels of by-catch) can be used to make inferences about population trends over time. In such cases, the goal ought to be to keep sources of bias (such as trackline detection probability) constant over time rather than minimizing bias. However, our study is novel in that it demonstrates that such platforms can allow researchers to estimate how many animals are in a given area. Estimates of absolute animal abundance are critical for, *inter alia*, calculating acceptable levels of anthropogenic mortality, or for quantifying predator needs before setting fishing quotas.

Variance of Abundance Estimates

Reliable estimation of the variance of abundance estimates from the spatial models is a difficult, and as yet unresolved, statistical challenge (Hedley et al. 1999, Forney 2000, Marques 2001, Bravington 2003). Using the same data collected in this study, Williams (2003) investigated four methods of estimating variance (bootstrapping vs. jackknifing, using Trip and Day as resampling units), resulting in 95% confidence intervals that differed by as much as an order of magnitude. In addition, the bootstrap estimates of variance showed substantial positive bias (i.e., the confidence limits did not contain the point estimate).

Given the competing interests of independence (resampling methods) and good spatial coverage (spatial modeling), the best compromise for this study was the jackknife estimator using each trip as the resampling unit. This provided the most reasonable estimates of variance. This finding echoes previous recommendations, which found the jackknife to perform well in a spatial modeling framework (Hedley 2000, Marques 2001). Clearly, further research is needed in this area, however, the confidence intervals we report represent a good first step. Although the reliability of our favored variance estimator is hindered by the small number of independent resampling units, the confidence intervals are likely to be unbiased, and they do incorporate uncertainty from all the primary sources of variation: detection function fitting; group size bias; and GAM-fitting. One area of ongoing research is to extract confidence intervals from the unexplained variance in the model-fitting process as a whole, but these methods will take some time to be developed and tested (Bravington 2002, 2003, 2004). In the meantime, we encourage researchers

who try these methods to collect as much data as possible.

In summary, these data indicate that the best abundance estimates currently available in the study area during the time of the survey were 1829 (95% CI: 978–3422) humpback whales, 4487 (95% CI: 1326–15 179) fin whales, and 1544 (95% CI: 1221–1953) minke whales.

Using Ships of Opportunity for Collecting and Modeling Line-transect Data

The key distinction between collecting data from a platform of opportunity and using an opportunistic data set is that in the former, the platform is opportunistic, but the researcher can ensure that study design and data collection are rigorous and dedicated. The same may or may not be true of study design and data collection in an existing, opportunistic data set. Researchers using free platforms must be prepared to be flexible, because what one gains financially and logistically, may be lost in terms of control over study design. However, in addition to using ships of opportunity for measuring distribution and abundance, line-transect data from non-randomized surveys can be very useful for traditional habitat modeling exercises.

The number of candidate environmental covariates in a spatial modeling framework is potentially large, and may include variables of direct relevance to whales, such as temperature, ice cover, and prey density. In addition to identifying how many animals are in an area, and how they are distributed, spatial modeling methods can describe associations to help us understand why whales are found where they are, or at least provide testable hypotheses about distribution. We strongly encourage the use of these platforms for collecting information on habitat preference and distribution of other marine wildlife, such as turtles, seabirds, and pinnipeds. However, in terms of estimating absolute abundance, we note that the detectability of these taxa is likely to be much lower than for baleen whales, whose blows are detectable on the scale of kilometers.

Synopsis

By incorporating information on the strip width effectively searched, our approach allows researchers to predict absolute abundance of wildlife in a study area, as well as preferred habitat within that area. Spatial modeling of line-transect data will allow surveys to be undertaken from ships of opportunity in regions where abundance and distribution data currently do not exist. Even abundance estimates on the correct order of magnitude may be of great value to researchers establishing conservation priorities in developing countries, and to managers in any country in the case of understudied taxa. As tourism expands globally to target currently untouched regions, it will be important to gain as much information from those platforms as possible.

GAM-based models enable estimation of the number of animals in a spatially flexible way, which will enable us to combine disparate surveys to detect changes in distribution and abundance over time. We encourage the use of the package *mgcv* for such methods, because it incorporates cross validation, and requires user-driven input in model selection. We welcome ongoing efforts to develop more rigorous variance estimators, but urge researchers to identify the most appropriate methods to quantify uncertainty in the meantime on a case-specific basis.

These data were collected, and some insight gained in the process, using minimal funds, and the volunteer efforts of people on a ship that was heading to Antarctica regardless of whether a scientist was on board. Other wilderness-based tourism platforms may be equally useful for conservation studies in understudied regions. Future work should consider the use of these ships to collect data that allow modeling of variables of biological relevance, namely those that influence animal distribution directly, rather than as habitat proxies. Habitat variables such as temperature, ice cover, salinity, and prey density can all be explored in a spatial modeling framework. When these data can be collected from a ship of opportunity, they should be. In a spatial modeling framework, line-transect data (i.e., effort, sightings, perpendicular distance, and relevant covariates) collected from free survey platforms become much more informative than sightings alone, and the techniques

are likely to have wide application to areas where financial resources are limiting and where population-level cetacean studies are just beginning.

Responses to this article can be read online at:
<http://www.ecologyandsociety.org/vol11/iss1/art1/responses/>

Acknowledgments:

Ship time and field support are gratefully acknowledged from Abercrombie & Kent, Cheeseman's Ecology Safaris, Fathom Expeditions, Marine Expeditions, the Mars family, and ships' staff and crew. Additional logistical and/or financial support was provided by McLean Foundation, Glendale Grizzly Trust, Jane Marcher Foundation, Newby Trust, Canadian Whale Institute, Cetacean Society International, Overseas Research Scholarship, Karten Foundation, and Whale and Dolphin Conservation Society. Data processing and analysis were aided by David Borchers, Steve Buckland, Louise Burt, Ari Friedlaender, Dave Johnston, Russell Leaper, Mike Lonergan, Charles Paxton, Val Smith, Len Thomas, Simon Wood, and Alexandre Zerbini. We thank Ian Boyd and Per Berggren, the Ecology & Society editorial staff, and an anonymous reviewer for helpful comments.

LITERATURE CITED

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267–281 in B. N. Petrov and F. Csaki, editors. 2nd International Symposium on Information Theory. AkadŠemiai Kiadi, Budapest, Hungary.

Baird, R. W., and H. Whitehead. 2000. Social organization of mammal-eating killer whales: group stability and dispersal patterns. *Canadian Journal of Zoology* 78:2096–2105.

Baker, C. S., and P. J. Clapham. 2002. Marine mammal exploitation: whales and whaling. Pages 446–450 in I. Douglas, editor. *Encyclopedia of global environmental change*. John Wiley, Chichester, UK.

Barlow, J., and S. Sexton. 1996. The effect of diving and searching behavior on the probability of detecting track-line groups, g_0 , of long-diving whales during line-transect surveys. NOAA National Marine Fisheries Service, Southwest Fisheries Center Administrative Report LJ-96-14.

Berggren, P., S. Brown, D. Gillespie, I. Kuklik, T. Lewis, J. Matthews, R. McLanaghan, A. Moscrop, and N. Tregenza. 2002a. Passive acoustic and visual survey of harbour porpoises (*Phocoena phocoena*) in Polish coastal waters confirms endangered status of Baltic population. Paper SC54/SM3 presented to the 55th meeting of the Scientific Committee of the International Whaling Commission (Shimonoseki, 2002). Shimonoseki, Japan. [Online.] URL: <http://www.iwcoffice.org>.

Berggren, P., P. R. Wade, J. Carlstrom, and A. J. Read. 2002b. Potential limits to anthropogenic mortality for harbour porpoises in the Baltic region. *Biological Conservation* 103:313–322.

Bigg, M A., P. F. Olesiuk, G. M. Ellis, J. K. B. Ford, and K. C. Balcomb. 1990. Social organization and genealogy of resident killer whales (*Orcinus orca*) in the coastal waters of British Columbia and Washington State. Pages 383–405 in P. S. Hammond, S. A. Mizroch, and G. P. Donovan, editors. *Individual recognition of cetaceans: use of photo-identification and other techniques to estimate population parameters*. Report of the International Whaling Commission (Special Issue 12).

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

Borchers, D. L., W. Zucchini, and R. M. Fewster. 1998b. Mark-recapture models for line transect surveys. *Biometrics* 54:1207–1220.

Borchers, D. L., and L. Burt. 2002. Generalized regression methods for estimating school size from line-transect data. Paper SC/54/IA23 presented to the 55th meeting of the Scientific Committee of the International Whaling Commission (Shimonoseki, 2002). Shimonoseki, Japan. [Online.] URL: <http://www.iwcoffice.org>.

- Bowen, W. D.** 1997. Role of marine mammals in aquatic ecosystems. *Marine Ecology Progress Series* **158**:267–274.
- Boyd, I. L.** 2002. Integrated environment–prey–predator interactions off South Georgia: implications for management of fisheries. *Aquatic Conservation: Marine and Freshwater Ecosystems* **12**:119–126.
- Braham, H. W., and M. E. Dahlheim.** 1982. Killer whales in Alaska documented in the platforms of opportunity program. *Reports of the International Whaling Commission* **32**:643–646.
- Bravington, M.** 2002. Spatial analyses of southern hemisphere minke whale data allowing for size bias and sightability. Paper SC/54/IA21 presented to the 54th meeting of the Scientific Committee of the International Whaling Commission (Shimonoseki, 2002). Shimonoseki, Japan. [Online.] URL: <http://www.iwcoffice.org>.
- Bravington, M.** 2003. On the misunderestimation of school size. Paper SC/55/IA10 presented to the 55th meeting of the Scientific Committee of the International Whaling Commission (Berlin, 2003). Berlin, Germany. [Online.] URL: <http://www.iwcoffice.org>.
- Bravington, M. V.** 2004. A new spatial abundance estimator, applied to simulated survey data. Paper SC/56/IA3 presented to the 56th meeting of the Scientific Committee of the International Whaling Commission (Sorrento, 2004). Sorrento, Italy. [Online.] URL: <http://www.iwcoffice.org>.
- Buckland, S., D. R. Anderson, K. P. Burnham, J. Laake, D. Borchers, and L. Thomas.** 2001. *Introduction to distance sampling: estimating abundance of biological populations*. Oxford University Press, Oxford, UK.
- Burnham, K. P., and D. R. Anderson.** 1976. Mathematical models for nonparametric inferences from line transect data. *Biometrics* **32**:325–336.
- Burns, W. C. G., and G. Wandesford-Smith.** 2002. The International Whaling Commission and the Future of Cetaceans in a Changing World. *RECIEL* **11**(2):199–210.
- Clarke, E. D., M. L. Burt, and D. Borchers.** 2000. Investigation of bias in GAM-based abundance estimation methods and their suitability for JARPA survey data. Paper SC/52/IA19 presented to the 52nd meeting of the Scientific Committee of the International Whaling Commission (Adelaide, 2000) Adelaide, Australia. [Online.] URL: <http://www.iwcoffice.org>.
- Dolphin, W. F.** 1987. Ventilation and dive patterns of humpback whales, *Megaptera novaenglia*, on their Alaskan feeding grounds. *Canadian Journal of Zoology* **65**:83–90.
- Donovan, G. P.** 1986. Behaviour of whales in relationship to management. *Reports of the International Whaling Commission (Special Issue 8)*. International Whaling Commission, Cambridge, UK.
- Efron, B.** 1979. Bootstrap methods: another look at the jackknife. *Annals of Statistics* **7**(1):1–26.
- Efron, B., and C. Stein.** 1981. The jackknife estimate of variance. *Annals of Statistics* **9**(3):586–596.
- Efron, B., and R. Tibshirani.** 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* **1**(1):54–75.
- Ford, J. K. B., G. M. Ellis, and K. C. Balcomb.** 2000. *Killer whales: the natural history and genealogy of *Orcinus orca* in British Columbia and Washington State*, 2nd edition. University of British Columbia Press, Vancouver, British Columbia, Canada.
- Forney, K.** 2000. Environmental models of cetacean abundance: reducing uncertainty in population trends. *Conservation Biology* **14**(5):1271–1286.
- Gowans, S., H. Whitehead, and S. K. Hooker.** 2001. Social organization in northern bottlenose whales, *Hyperoodon ampullatus*: not driven by deep-water foraging? *Animal Behaviour* **62**:369–377.
- Gregg, E. J., L. Nichol, J. K. B. Ford, G. Ellis, and A. W. Trites.** 2000. Migration and population structure of northeastern Pacific whales off coastal British Columbia: an analysis of commercial whaling records from 1908–1967. *Marine Mammal Science* **16**:699–727.

- Gregr, E. J., and A. W. Trites.** 2001. Predictions of critical habitat for five whale species in the waters of coastal British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences* **58**:1265–1285.
- Hammond, P. S., P. Berggren, H. Benke, D. L. Borchers, A. Collet, M. P. Heide-Jørgensen, S. Heimlich, A. R. Hiby, M. F. Leopold, and N. Øien.** 2002. Abundance of harbour porpoise and other cetaceans in the North Sea and adjacent waters. *Journal of Applied Ecology* **39**:361–376.
- Hastie, T. J., and R. Tibshirani.** 1990. *Generalized additive models*. Chapman and Hall, London, UK.
- Hedley, S. L.** 2000. *Modelling heterogeneity in cetacean surveys*. Dissertation, University of St. Andrews, St. Andrews, UK.
- Hedley, S. L., S. T. Buckland, and D. L. Borchers.** 1999. Spatial modelling from line transect data. *Journal of Cetacean Resource Management* **1** (3):255–264.
- Hedley, S., S. Reilly, J. Borberg, R. Holland, R. Hewitt, J. Watkins, M. Naganobu, and V. Sushin.** 2001. Modelling whale distribution: a preliminary analysis of data collected on the CCAMLR-IWC Krill Synoptic Survey, 2000. Paper SC/53/E9 presented to the 54th meeting of the Scientific Committee of the International Whaling Commission (London, 2001). London, UK. [Online.] URL: <http://www.iwcoffice.org>.
- Hooker, S. K., H. Whitehead, and S. Gowans.** 1999. Marine protected area design and the spatial and temporal distribution of cetaceans in a submarine canyon. *Conservation Biology* **13** (3):592–602.
- Hoyt, E.** 1997. *The potential of whale watching in Europe*. Whale and Dolphin Conservation Society, Bath, UK.
- International Association of Antarctic Tour Operators.** 2000. Overview of Antarctic tourism, 2000. *Information Paper 33 presented to the Antarctic Treaty XII Special Consultative Meeting*. The Hague, Netherlands. [Online.] URL: http://www.iaato.org/xxii_iaato_over2000.html.
- International Whaling Commission (IWC).** 2005. Annex E. Report of the Standing Working Group on the Development of an Aboriginal Subsistence Whaling Management Procedure (AWMP). *Journal of Cetacean Research Management (Suppl.)* **7**. [Online.] URL: <http://www.iwcoffice.org>.
- Jahoda, M., C. L. Lafortuna, N. Biassoni, C. Almirante, A. Azzellino, S. Panigada, M. Zanardelli, and G. Nortarbartolo di Sciara.** 2003. Mediterranean fin whale's (*Balaenoptera physalus*) response to small vessels and biopsy sampling assessed through passive tracking and timing of respiration. *Maritime Mammal Science* **19** (1):96–110.
- Laake, J. L.** 2001. Excel geometry functions. [Online.] URL: <http://nmml.afsc.noaa.gov/Software/ExcelGeoFunctions/excelgeofunc.htm>.
- Laws, R. M.** 1977. Seals and whales of the Southern Ocean. *Philosophical Transactions of the Royal Society of London Series B* **279**:81–96.
- Lindenmayer, D. B., R. B. Cunningham, C. MacGregor, R. D. Incoll, and D. Michael.** 2003. A survey design for monitoring the abundance of arboreal marsupials in the Central Highlands of Victoria. *Biological Conservation* **110**:161–167.
- Marques, F. F. C.** 2001. *Estimating wildlife distribution and abundance from line transect surveys conducted from platforms of opportunity*. Dissertation, University of St. Andrews, St. Andrews, UK.
- Marsh, H., and D. F. Sinclair.** 1989. Correcting for visibility bias in strip transect aerial surveys of aquatic fauna. *Journal of Wildlife Management* **53**:1017–1024.
- McCullagh, P., and J. A. Nelder.** 1989. *Generalized linear models*. Chapman and Hall, London, UK.
- Miller, R. G.** 1974. The jackknife—a review. *Biometrika* **61**(1):1–15.
- Mizroch, S. A.** 1983. *Reproductive rates in southern hemisphere baleen whales*. Dissertation, University of Washington, Seattle, Washington, USA.
- Mizroch, S. A.** 1984. The development of balaenopterid whaling in the Antarctic. *Cetus* **5** (2):6–10.

Moore, M. J., S. D. Berrow, B. A. Jensen, P. Carr, R. Sears, V. J. Rowntree, R. Payne, and P. K. Hamilton. 1999. Relative abundance of large whales around South Georgia (1979–1998). *Marine Mammal Science* 15(4):1287–1302.

Nelson, J. G. 1994. The spread of ecotourism—some planning implications. *Environmental Conservation* 21(3):248–255.

Northridge, S.P., M. L. Tasker, A. Webb, and J. M. Williams. 1995. Distribution and relative abundance of harbour porpoises (*Phocoena phocoena*, L.), whitebeaked dolphins (*Lagenorhynchus albirostris*, Gray) and minke whales (*Balaenoptera acutorostrata*, Lacepede) around the British Isles. *ICES Journal of Marine Science* 52:55–66.

Northridge, S. P., and R. J. Hofman. 1999. Marine mammal interactions with fisheries. Pages 99–119 in J. R. Twiss, Jr. and R. R. Reeves, editors. *Conservation and management of marine mammals*. Smithsonian Institution, Washington, D. C., USA.

R Development Core Team. 2003. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. [Online.] URL: <http://www.R-project.org>.

Strindberg, S. 2001. *Optimized automated survey design in wildlife population assessment*. Dissertation, University of St. Andrews, St. Andrews, UK.

Talbot, L. 1974. The great whales and the International Whaling Commission. Pages 232–236 in J. McIntyre, editor. *Mind in the waters*. McClelland and Stewart, Toronto, Ontario, Canada.

Thomas, L., J. L. Laake, S. Strindberg, F. F. C. Marques, S. T. Buckland, D. L. Borchers, D. R. Anderson, K. P. Burnham, S. L. Hedley, and J. H. Pollard. 2002. *Distance 4.0. Release 2*. Research Unit for Wildlife Population Assessment, University of St. Andrews, St. Andrews, UK. [Online.] URL: <http://www.ruwpa.st-and.ac.uk/distance/>.

Trites, A. W., V. Christensen, and D. Pauly. 1997. Competition between fisheries and marine mammals for prey and primary production in the Pacific Ocean. *Journal of Northern Atlantic*

Fisheries Science 22:173–187.

Vidal, O. 1993. Aquatic mammal conservation in latin america: problems and perspectives. *Conservation Biology* 7(4):788–795.

Wade, P. R. 1998. Calculating limits to the allowable human-caused mortality of cetaceans and pinnipeds. *Marine Mammal Science* 14(1):1–37.

Williams, R. 2003. *Cetacean studies using platforms of opportunity*. Dissertation, University of St. Andrews, St. Andrews, UK.

Wood, S. N. 2000. Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society, Series B* 62:413–428.

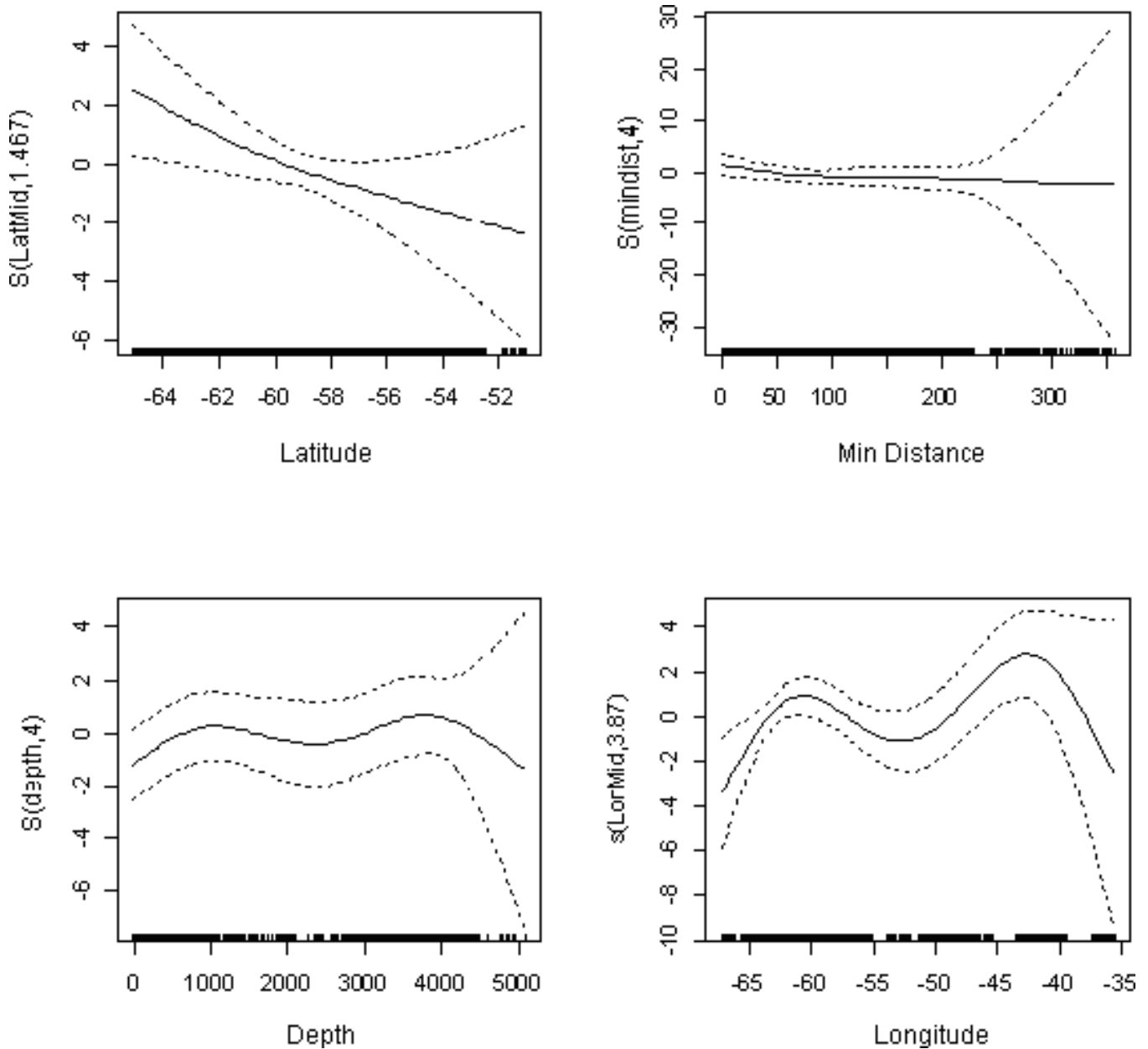
Wood, S. N. 2001. mgcv: GAMs and Generalized Ridge Regression for R. *R News* 1(2):20–25.

Wood, S. N. 2003. Thin plate regression splines. *Journal of the Royal Statistical Society, Series B* 65(1):95–114.

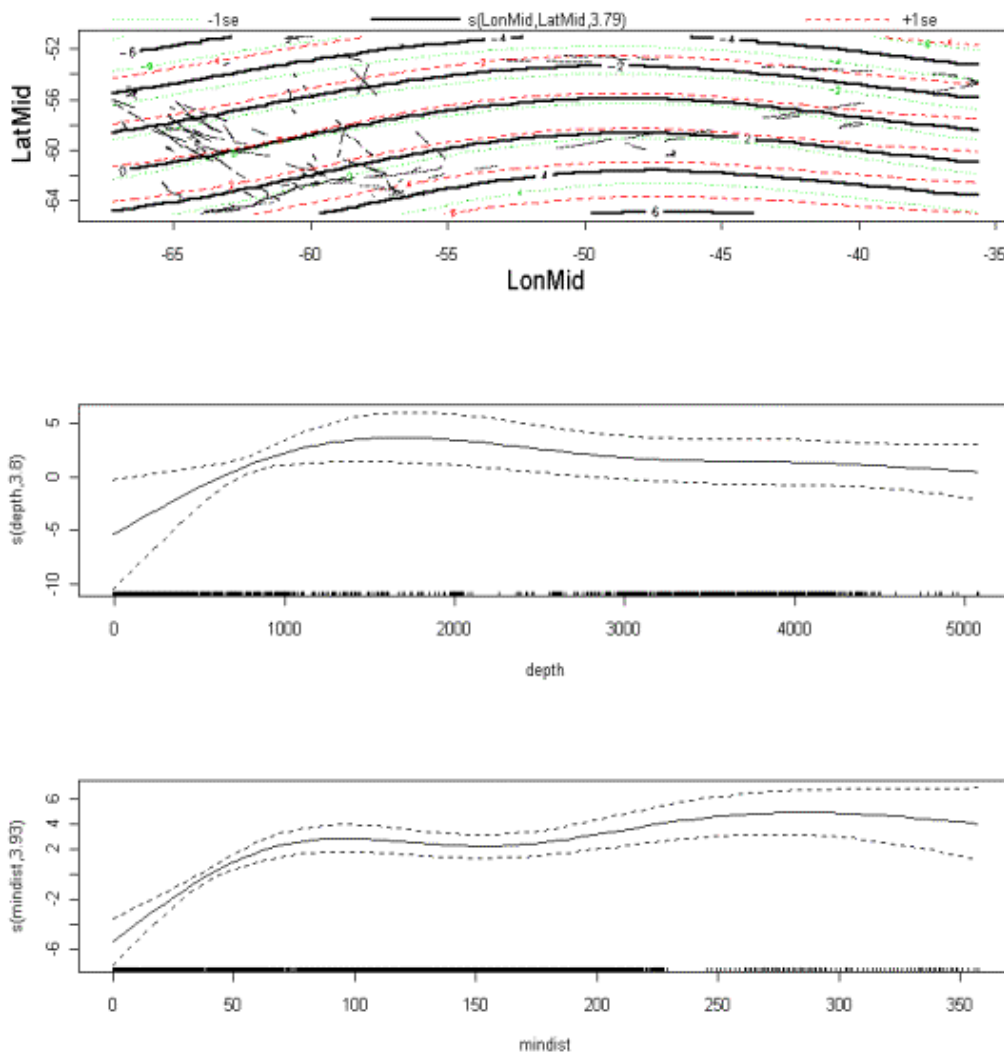
Appendix 1. Expanded Methods Section.

[Please click here to download file 'appendix1.pdf'.](#)

APPENDIX 2. Smoothed components (solid line) of four explanatory variables (x-axes) used in the fitted GAM and the response variable, density of humpback whale schools. Each explanatory variable was allowed up to 9 df, and degree of smoothing was automated by mgcv. Each x-axis contains a rugplot, where small ticks mark locations of observations. Regions of high whale density are identified by values above zero on the y-axis. Note the different scale of each y-axis, which is labeled s(covariate name, estimated degrees of freedom). The dashed lines represent ± 2 standard errors, or roughly 95% confidence intervals.



APPENDIX 3. One two-dimensional (LonMid, LatMid) and two one-dimensional (depth + mindist) smoothing splines (solid line) of the fitted GAM (x-axes) and the response variable, density of fin whale schools. Each explanatory variable was allowed up to 9 df, and degree of smoothing was automated by mgcv. Each x-axis contains a rugplot, where small ticks mark observed values. Regions of high whale density are identified by values above zero on the y-axis. Note the different scale of each y-axis, which is labeled s(covariate name, estimated degrees of freedom). The dashed lines represent ± 2 standard errors, or roughly 95% confidence intervals.



APPENDIX 4. Smoothed components (solid line) of three explanatory variables (x-axes) used in the fitted GAM and the response variable, density of minke whale schools. Each explanatory variable was allowed up to 9 df, and degree of smoothing was automated by mgcv. Each x-axis contains a rugplot, where small ticks mark locations of observations. Regions of high whale density are identified by values above zero on the y-axis. Note the different scale of each y-axis, which is labeled $s(\text{covariate name, estimated degrees of freedom})$. The dashed lines represent ± 2 standard errors, or roughly 95% confidence intervals.

