

Simulations of IDCR and SOWER Antarctic minke whale abundance surveys

DEBRA L. PALKA¹ AND DANIEL W. SMITH²

Contact email: debra.palka@noaa.gov

ABSTRACT

The International Whaling Commission (IWC) conducted abundance line-transect surveys of Antarctic minke whales (*Balaenoptera bonaerensis*) during the International Decade of Cetacean Research (IDCR) between 1978/79 and 2003/04. These surveys continued throughout 2004/05–2009/10 under the Southern Ocean Whale and Ecosystem Research (SOWER) programme. Several methods used to analyse these data resulted in very different abundance estimates. In order to evaluate the accuracy of these methods, the IWC Scientific Committee recommended using simulated line-transect data. The software package TRANSIM was therefore modified to generate simulated line-transect data that emulated an IDCR/SOWER survey in a study area populated with whales that followed the distribution, density and behaviour of Antarctic minke whales. Simulations incorporated factors that could potentially bias resulting abundance estimates, and in order to investigate the effects of these factors on the resulting abundance estimates, simulated datasets were generated for different scenarios, each of which incorporated combinations of factors that were chosen using the mirror-image fold-over partial factorial method that is designed to identify factors related to biased estimates. This paper describes the programme used to generate these datasets and reports on the robustness of the statistical methods applied to analyse them. In addition, a brief discussion is included on how this information was used when the actual line-transect data were analysed. The statistical methods investigated were the hazard probability method (OK), integrated hazard probability method (IM), spatial point independence method (SPL) and standard method. After developers analysed the simulated data, they used this information to improve their analytical methods. The results reported here focus on the performance of the most advanced versions of the OK and SPL methods compared with the standard method as of 2009. In general, SPL estimates of group density were negatively biased, OK estimates were positively biased, standard estimates were the most negatively biased and IM estimates were the least biased. Density estimates from SPL were less biased than estimates from OK in the complex scenarios that contained spatial gradients and correlations between density, location and weather conditions. The OK method was less robust when: (1) closing-mode data collection procedures were simulated; (2) whale groups were diving in a non-synchronous fashion; and (3) there were complex spatial patterns. In almost every case, the density estimates from the OK method were greater than the SPL methods, on average by 10–15%. These observations assisted the development of more robust analysis methods and highlighted which factors should be addressed when analysing the actual data.

KEYWORDS: ABUNDANCE ESTIMATE; G(0); MONITORING; SCHOOL SIZE; SURVEY-VESSEL; ANTARCTIC; SIMULATED DATA; MIRROR-IMAGE FOLD-OVER PARTIAL FACTORIAL DESIGN

¹ National Marine Fisheries Science Centre, Woods Hole, MA 02543, USA.

² Integrated Statistics, Woods Hole, MA 02543, USA.

INTRODUCTION

Line-transect methodology is a conventional means of estimating whale abundance and has been applied to many whale species. Skaug & Schweder (1999) demonstrate that heterogeneity in factors such as weather and whale surfacing-diving behaviour could lead to severe underestimation of abundance from line-transect surveys. Consequently, inclusion of such factors is required when the goal is to calculate a precise and unbiased estimate of absolute abundance. However, the best way to account for these types of factors can be difficult to determine. The use of simulated data sets is one way to ascertain whether an analysis method is robust and results in a precise and unbiased abundance estimate.

In general, simulated datasets can serve as either an evaluation or development tool. For example, scientists from the IWC's Scientific Committee used simulated datasets as an evaluation tool to investigate the performance of analytical methods used to estimate abundance of North Atlantic minke whales (*Balaenoptera acutorostrata*) around Norway. They compared the true density of the simulated data to the estimated density resulting from applying the analysis methods to the simulated data. A similar type of comparison was performed to evaluate four analytical methods that did not require the assumption that $g(0)$ equals one (Polacheck *et al.*, 2000), where $g(0)$ is the probability of detecting a group on the track line.

In 2002, the Scientific Committee recommended using simulated datasets to evaluate the robustness of analytical methods that might be used to estimate the abundance of the Antarctic minke whale (*Balaenoptera bonaerensis*) from line-transect data. These data were collected under the IWC's IDCR and SOWER programmes which conducted three sets of circumpolar (CP) surveys around the Antarctic: CPI spanned the austral summers of 1978/79–1983/84; CPII spanned 1984/85–1990/91; CPIII spanned 1991/92–2003/04. For this evaluation, simulated datasets served as a development tool to refine analytical methods to account for the specifics of these surveys. The ultimate goal was to improve the estimates of abundance when applying the analysis method(s) to the actual data. To this end, simulated data were generated that explicitly incorporated heterogeneity in factors that could bias the results when analysed using line-transect analytical methods. The simulation data attempted to mimic data collected using procedures specific to the IDCR and SOWER surveys, in addition to biological characteristics, such as distribution, density and behaviour of Antarctic minke whales. Four methods were used to analyse the simulated data: (1) the hazard probability (OK) method (Okamura & Kitakado, 2010); (2) the integrated hazard probability (IM) method (Cooke, 2002); (3) the spatial point independence (SPL) method (Bravington & Hedley, 2010); and (4) the standard analysis method (Branch, 2006).

This paper describes the model characteristics used to create the simulated data, potential biasing factors that were included, methods used to investigate the robustness of the analysis methods and the resulting comparisons between true and estimated values for each analysis method. There is also a brief discussion of how this information on the robustness of these methods was used when the actual line-transect data were analysed.

METHODS

Software was developed to generate simulated data mimicking IDCR and SOWER line-transect sighting surveys in the Antarctic. The Scientific Committee suggested possible biasing factors to include. In order to investigate the main effects of multiple factors while minimising the effects of two-way and higher interactions, a mirror-image fold-over partial factorial design (Petersen, 1985; NIST/SEMATECH, 2003) was used to determine which factors were present in sets of scenarios. The advantage of this design is that a carefully chosen subset of all permutations can be used to expose information about the most important features while using only a fraction of the effort. Each scenario had a different combination of five to seven factors that could result in biased density estimates from line-transect analyses. The developers of analytical methods analysed the simulated datasets in a blind experiment: they did not know the values of factors involved or the model parameters. Using analysis of variance (ANOVA) methods, the resulting estimated density values, mean group size and density of groups were compared with their true values to identify less robust aspects of the analysis methods. The developers then further refined these aspects and re-analysed the simulated data.

Simulated data

The software used to create the simulated line-transect data was an updated version of TRANSIM, a software package written in C and C++ (Polacheck *et al.*, 1996). The basic simulation model consisted of a ‘universe’ of surfacing/diving whales in a two-dimensional plane (the ocean), with a ship surveying for 16–22 hours per day while traveling at 11.5 knots (when on the track line). The Antarctic ice edge was assumed to be immediately south of the area surveyed by the ship. Each day the ship either started near the ice edge and travelled north away from the ice or started at the most northerly end of a track line and travelled south towards the ice edge. Depending on the scenario, the true distribution of whale groups was either uniform or clustered, could follow spatial gradients, or any combination of these (e.g., Fig. 1). The distribution of group sizes was derived either from a theoretical Poisson distribution that could follow spatial gradients or from empirically measured group sizes. In cases where group sizes were greater than one, all animals in a group could surface either simultaneously or in a non-synchronised fashion. In the dive model, the duration of a surfacing event between dives was one second (one time step in the model). In the case of synchronised diving, the mean dive time for a group was 75 seconds, with a range of five to 480 seconds (Fig. 2) (Ward, 1988). The software also allowed whales to respond to the ship in a user-specified manner, though this feature was not used in this specific investigation (see Appendix 1 for more information).

The simulated line-transect survey followed the IDCR and SOWER data collection protocols. Specifically, there was one ship with teams of observers on three observation platforms. The three platforms were: (1) the main observation platform, called the Topman team; (2) an additional team, called the IO team, that surveyed only during the independent observer (IO) mode; (3) the Bridge team, located on the bridge, which recorded all the data from the three teams and searched for whales. All teams searched the area in front of the ship and every instance of surfacing within a user-specified area was tested for whether it was detected by each team. The probability of detecting a whale group surfacing was a function of a variety of factors, including the relative position and heading

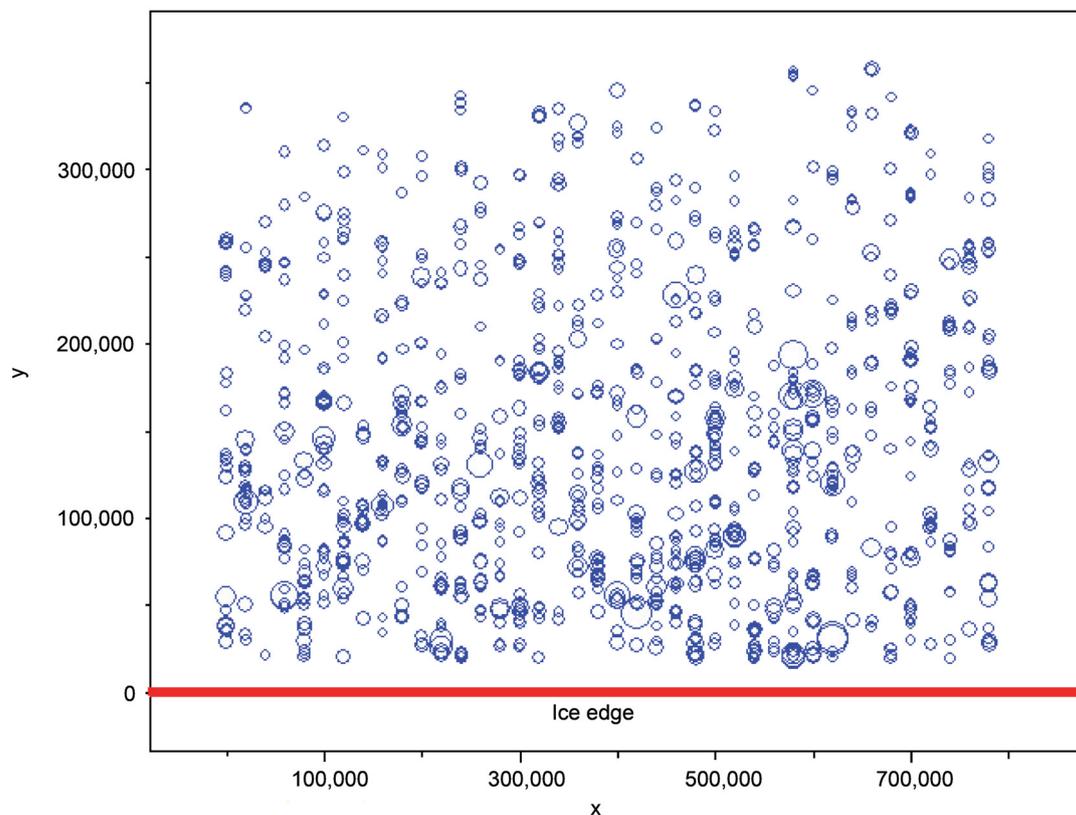


Figure 1. True locations of sightings for a scenario (sc09) with clustered groups (groups more tightly clumped together near the ice edge), a gradient in density (higher group density near the ice edge) and a gradient in the size of groups (larger groups near the ice edge). Each circle represents a group of animals where the size of the circle is proportional to its group size.

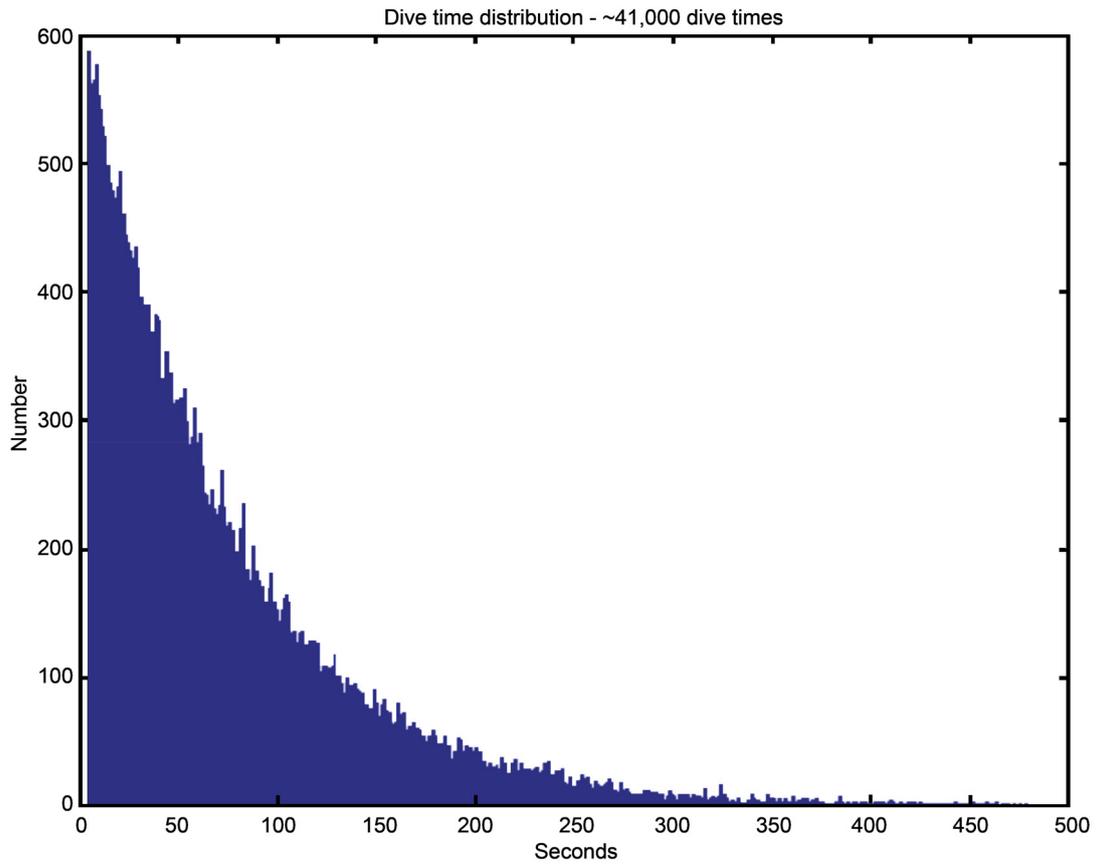


Figure 2. Dive time distribution for groups of whales diving in synchrony.

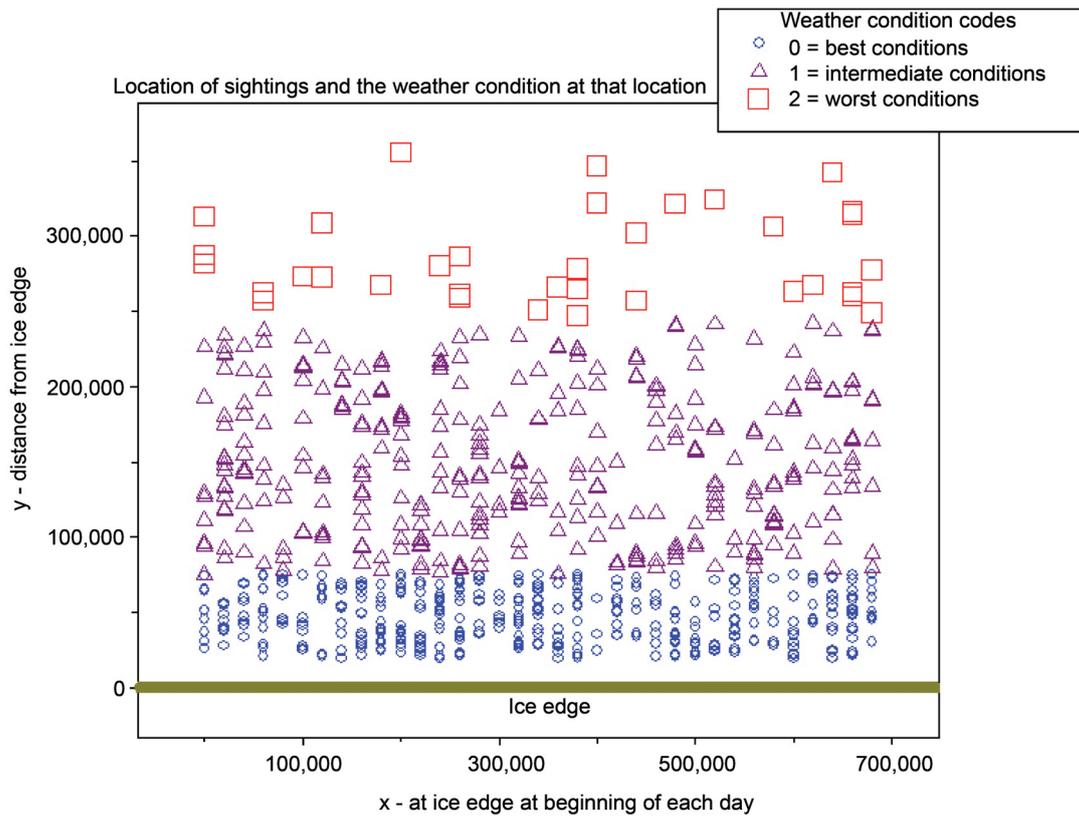


Figure 3. True locations of sightings detected in one replicate. There was a group density gradient (more groups near the ice edge), a weather condition gradient (better weather (code = 0) near the ice edge) and the probability of detecting a group was dependent on the weather condition at that location (the worse the weather, the lower the probability of detecting a group).

of the whale group, weather conditions, group size, initial cue (body or blow), observer team within a platform, and the platform (e.g., Fig. 3). See Appendix 1 for more information on the detection function models, parameters and example plots. The survey protocol followed either the IO-mode procedure, the Closing-mode procedure or a mixture of these. During the IO-mode procedure, the ship never left the track line, which means estimates of group sizes were most likely biased low because there was insufficient time to determine the true number of whales in the group. During the Closing mode procedure, when a whale group was detected within a user-specified range around the ship, surveying effort stopped, the ship left the track line to approach the group and confirm the species identification and group size, then the ship returned to the track line and surveying resumed. The software could also add errors to the recorded values of the timing of a sighting, radial distance, sighting angle and group size.

Scenarios

A mirror-image fold-over partial factorial design was used to investigate the main effect of each separate factor when multiple factors were operating simultaneously. An advantage of this design is that the main effects are free from two-way and higher interactions, so that the effects are more straightforward to interpret. This type of design led to sets of 16 scenarios (sc), where each scenario included a different combination of factors. These sets of simulated scenarios (Table 1) were created in 2004 (sc01-16), 2005 (sc17-32) and 2008 (sc39-54). In 2006–07, the programme code was upgraded to eliminate any slight chance of error by explicitly assigning all variables an initial value rather than using the default initialisation. In addition, we generated several more scenarios (sc33-38) that included more complex combinations of factors. In 2012–13, one additional scenario

Table 1
Factors included in each set of scenarios.

2004 (sc01-16)	
1.	Whale density followed a horizontal (along ice) spatial gradient
2.	Group size followed a horizontal (away from ice) spatial gradient
3.	Whale groups were clustered
4.	Survey data were collected either under IO mode or alternating between IO and Closing modes
5.	Probability of detecting a group was dependent on group size
6.	Probability of detecting a group was dependent on weather conditions
7.	Probability of detecting a group was dependent on the initial cue
2005 (sc17-32)	
1.	Recorded data included group size errors
2.	Whale density was correlated with weather-related sighting conditions
3.	Whale density followed a complex combination of vertical and horizontal density gradients
4.	Survey data were collected either under IO mode or alternating between IO and Closing modes
5.	Individual whales within a group surfaced in either a synchronised or non-synchronised fashion
2006–07 (sc33-38)	
1.	Recorded data included group size errors greater than in 2005
2.	Probability of detecting a group was dependent on weather-related sighting conditions
3.	Whale density followed a gradient either monotonic or non-monotonic (e.g., Fig. 4)
4.	Whale density followed a combination of vertical and horizontal density gradients
5.	Group size followed a spatial gradient
6.	Individual whales within a group surfaced in either a synchronised or non-synchronised fashion, similar to 2004 scenarios
7.	Recorded data included errors in radial distance and sighting angles
2008 (sc39-54)	
1.	Recorded data included measurement errors in time, radial distance, angle and group size
2.	Whale density followed a non-linear horizontal gradient
3.	Some groups were incorrectly assigned as a duplicate sighting
4.	Individual whales within a group surfaced in either a synchronised or non-synchronised fashion, similar to earlier scenarios
5.	Location correlation was present; i.e., simultaneous group size gradient (larger groups near ice), a non-linear vertical density gradient (higher density near ice) and a weather gradient (better weather near ice), with the detection function dependent on group size and weather
2012-03 (sc55)	
1.	Detection function dependent on radial distance, sighting angle, group size and weather conditions
2.	Non-monotonic gradient in whale group density and no clustering of whale groups
3.	Linear gradient in group size distribution (more large groups near ice edge)
4.	Group size distribution derived from empirical distribution (different from sc01-54)
5.	Diving/surfacing pattern derived from empirical distribution (different from sc01-54)
6.	Individual whales within a group surfaced only in a non-synchronised fashion
7.	Detection functions more similar to the recent empirical distributions only and the Bridge observers now recorded sightings seen at same time as Topman observers (different from sc01-54)
8.	Survey data were collected only under the procedure of alternating between IO and Closing modes
9.	Recorded data included errors in radial distance, sighting angle and group size (different from sc01-54)

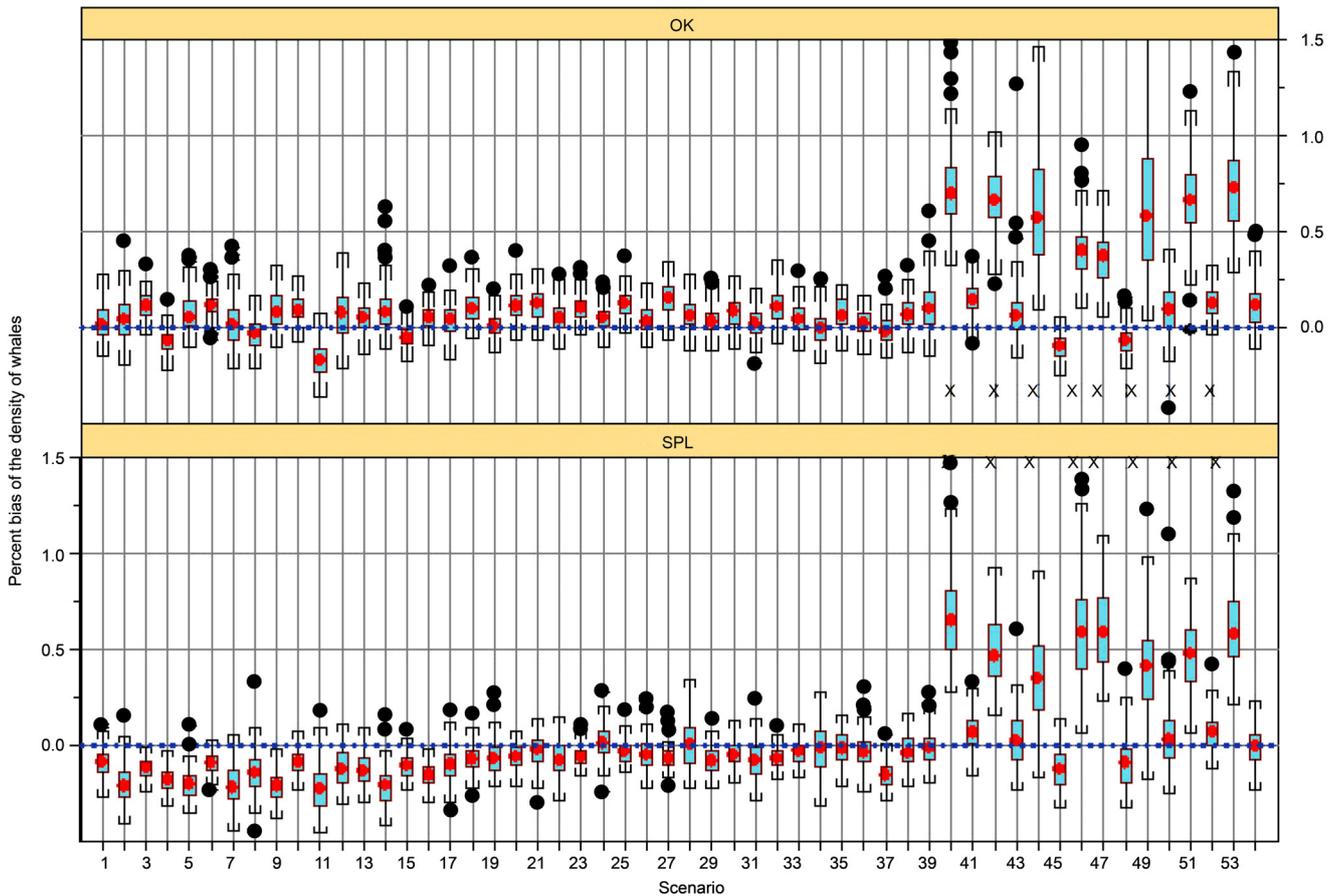


Figure 4. Box plots of the percent mean relative bias ($100 \cdot (\text{mean} - \text{true}) / \text{true}$) of the density of whales from the 100 iterations within each scenario for the 2009 versions of the OK and SPL methods. Scenarios marked with 'x' are those with misidentified duplicate errors.

(sc55) was developed that simultaneously had almost all the possible major biasing factors, plus diving patterns and group size bias derived from empirical data collected during an IDCR/SOWER survey. One hundred replicates were generated for each scenario to investigate the variability of the results. Each replicate contained the same factors and model parameterisations but with random variations within these constraints.

Investigation of effects

Estimates of three variables were provided for each replication by each analysis method: the density of individual whales, the density of groups of whales and the average group size. In order to investigate the effects of potential biasing factors on each of the three estimates, each annual set of 16 scenarios was analysed independently. For each estimate, the percent relative bias, $100 \cdot (\text{estimate} - \text{true}) / \text{true}$, was calculated for each replicate. A scenario was summarised by the percent mean relative bias, $100 \cdot (\text{mean} - \text{true}) / \text{true}$, where the mean was the average of the 100 replicates within a scenario. A set of scenarios was summarised by the mean, median and root mean square of the percent mean relative biases of the scenarios within a set.

An ANOVA was used to quantify the main effects and identify which factors corresponded with most biased estimates. For example, the ANOVA for the 2004 set of scenarios was:

$$\text{Percent mean relative bias}_{2004} \sim \text{vertical density gradient} + \text{size gradient} + \text{clustered} + \text{mode change} + \text{detect}_{\text{size}} + \text{detect}_{\text{weather}} + \text{detect}_{\text{cue}}$$

Where each factor on the right-hand side of the equation was a dummy variable for each scenario in the set: -1 indicated a factor was 'absent' or +1 indicated a factor was 'present'.

Similarly, the ANOVA models for the 2005 and 2008 scenarios were, respectively:

$$\text{Percent mean relative bias}_{2005} \sim \text{size errors} + \text{weather density} + \text{vertical and horizontal density} + \text{mode change} + \text{dive pattern}$$

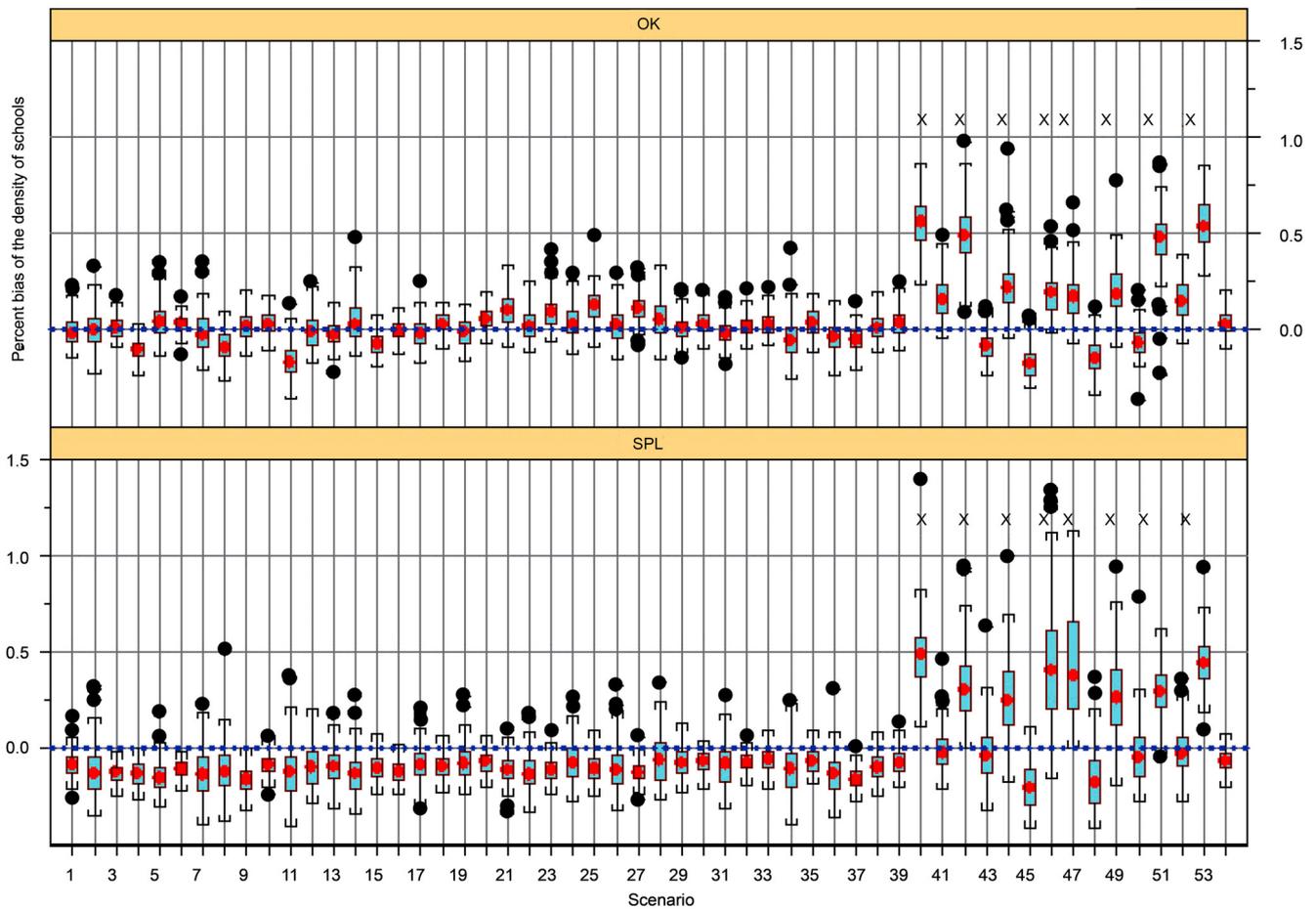


Figure 5. Box plots of the percent mean relative bias ($100 \cdot (\text{mean} - \text{true}) / \text{true}$) of the density of groups from the 100 iterations within each scenario for the 2009 versions of the OK and SPL methods. Scenarios marked with 'x' are those with misidentified duplicate errors.

and:

$$\text{Percent mean relative bias}_{2008} \sim \text{measurement errors} + \text{horizontal density gradient} + \text{mis-ID groups} \\ + \text{dive pattern} + \text{location correlation}$$

Table 2 provides the definition of 'present' and 'absent' for each set of scenarios.

RESULTS

Comparison to the true values

The true density values of the simulated data ranged between 0.0295–0.1066 whales/km² and 0.0288–0.0454 groups of whales/km², while the true mean group size ranged between 1.91–3.51 (Tables 3–7). Several different versions of the OK and SPL analysis methods analysed the simulated data. The ANOVA factor effects revealed the level of robustness of the different versions (Tables 8–10). The values of the percent mean relative bias (Tables 3–7) and the robustness patterns described below (Figs. 4–9) only focus on the most recently developed and best versions of each analysis method: OK09 and SPL09. No developer analysed the 2012–13 scenario (sc55), but its parameterisations are included for completeness and remain available to any potential analyst.

Comparing the OK09, SPL09 and standard methods across all scenarios led to the following general observations (Tables 3–7):

- Across all scenarios, the estimates of mean group sizes from all three methods were similar. They were relatively unbiased, with the average (and median) percent mean relative bias ranging within 2–6%. The OK09 and SPL09 methods had similar levels of variability (about RMS = 6.5), while the standard method was the most variable (RMS = 9.7).

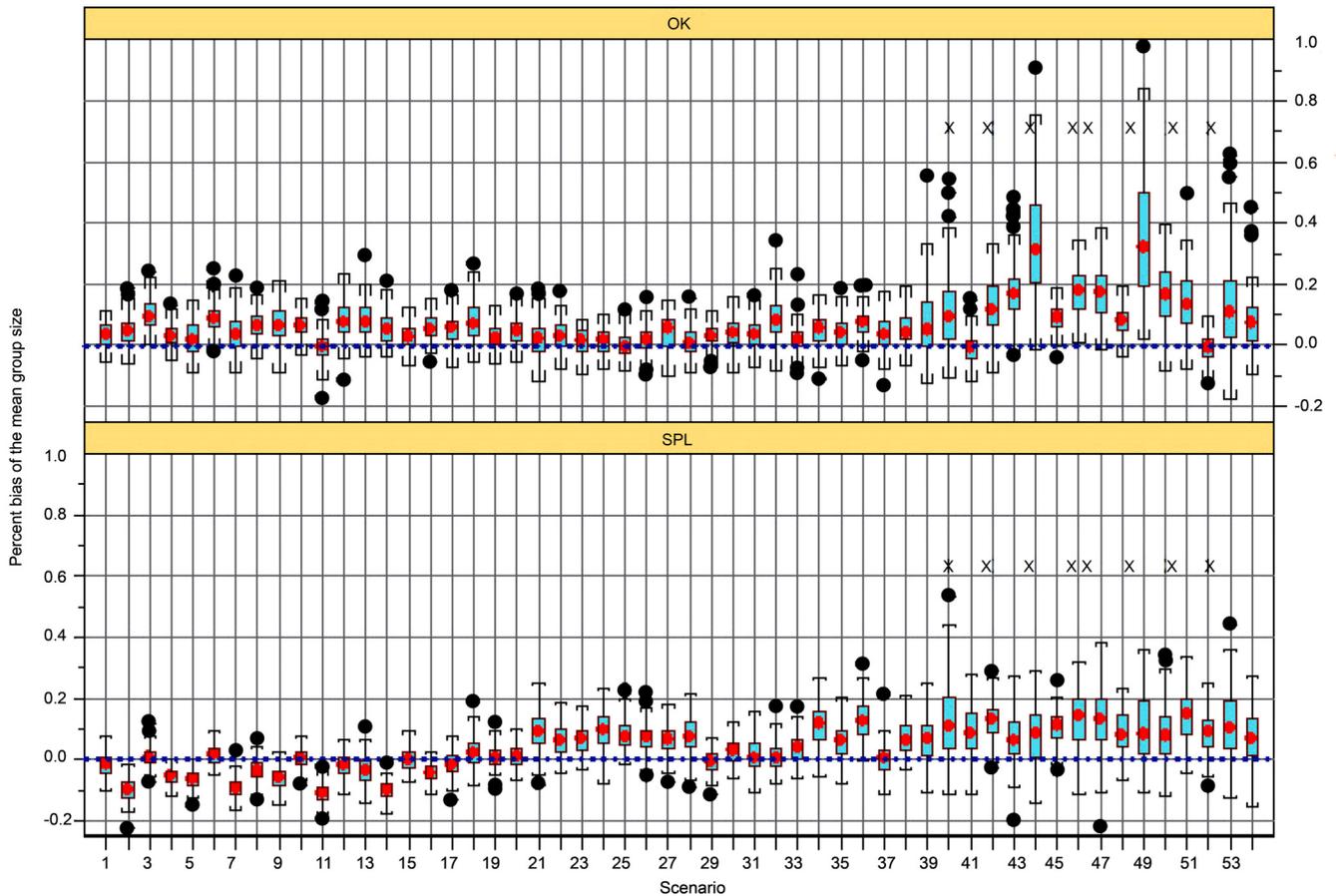


Figure 6. Box plots of the percent mean relative bias ($100 \cdot (\text{mean} - \text{true}) / \text{true}$) of the mean group sizes from the 100 iterations within each scenario for the 2009 versions of the OK and SPL methods. Scenarios marked with 'x' are those with misidentified duplicate errors.

- Across all scenarios, the estimate of group density from the SPL09 method was nearly always negatively biased (average percent mean relative bias = -9.5% , median = -9.6% , RMS = 10.3). In contrast, the group density from the OK09 method was positively biased to a small extent and slightly less variable (average percent mean relative bias = 0.8% , median = 2.2% , RMS = 7.2). The standard method was the most negatively biased and most variable (average percent mean relative bias = -19.6% , median = -19.7% , RMS = 21.4).
- The estimates of whale density for all three methods followed the same patterns as the group density estimates.

The levels of bias in the results were different when focusing only on the complex and more realistic scenarios that incorporated many of the potentially biasing factors (measurement error, spatial patterns, correlations between weather and animal density, and non-synchronised diving), but not misidentification of duplicate errors (Table 6). Specifically, the estimates of whale density from the SPL09 method were less biased (average percent mean relative bias = 0.1, median = 2.0, RMS = 6.9) than the OK09 method (average percent mean relative bias = 6.8, median = 10.8, RMS = 11.1). Both OK09 and SPL09 were less biased and less variable than the standard method (average percent mean relative bias = -10.9 , median = -11.8 , RMS = 17.0). The standard method resulted in less biased results when analysing complex scenarios than less complex scenarios, as described above.

Examining distributions of the relative bias for each iteration within a scenario for the OK09 and SPL09 analysis methods led to the following observations (Figs. 4–6). As a simple indicator, we used the term 'not biased' to describe situations where the true value was within the centre half of the distribution of estimates (25–75 percentiles; blue boxes in box plots).

Table 2
Definition of absent and present for each factor included in the sets of scenarios.

Year	Factor	Absent	Present
All	Mode change	Only IO mode	Closing and IO modes alternated
2004	Vertical (away from ice) density gradient	Groups randomly distributed	Group density highest at the ice edge and linearly decreased with distance from ice (Fig. 1)
2004	Size gradient	Group sizes randomly distributed	Group sizes largest at the ice edge and linearly decreased with distance from ice edge
2004	Clustered groups	Groups randomly distributed	Groups were spatially clustered
2004	DF size	Detection function not dependent on group size	Detection function dependent on group size
2004	DF weather	Detection function not dependent on weather conditions	Detection function dependent on weather conditions
2004	DF cue	Detection function not dependent on the initial cue	Detection function dependent on the initial cue
2005	Size error	No measurement error in the recorded group size	Measurement error added to the recorded sizes of groups that were not closed upon
2005	Weather-Density	No correlation between weather and true whale density	True whale density correlated with weather condition at the time of sighting
2005	Vertical and horizontal density	Groups randomly distributed	Whale density followed a gradient in both the vertical (away from ice) and horizontal (along ice) direction
2005, 2008	Non-synchronised diving	Individuals within a group all dove in synchrony	Individuals within a group followed different dive patterns, so the group did not dive in synchrony
2008	Measurement error	All values recorded without error	Measurement error added to recorded values of time, radial distance, angle and group size
2008	Horizontal density gradient	Groups randomly distributed	Whale density followed a horizontal (along ice) gradient
2008	Misidentified some groups	All recorded duplicates recorded correctly	Some duplicate sightings misidentified
2008	Location correlation	No location correlations (i.e., groups, group sizes and weather conditions not related to location)	Whale density followed a non-linear vertical gradient; group sizes followed a horizontal gradient; weather conditions changed over the day and influenced the probability of detecting a group

In general, estimates from SPL09 were less biased than estimates from OK09 in the complex scenarios (sc39-54). In contrast, OK09 was less biased in the less complex scenarios (sc01-38).

In the complex scenarios (sc39-54) without misidentification of duplicates, the estimates of density of groups from the OK09 method were generally negatively biased and the estimates of group size were positively biased, resulting in a mixture of positively, negatively or not biased estimates of whale density. The SPL09 method also followed this pattern, but the level of bias was smaller, i.e., the estimates of group size were generally slightly positively biased while estimates of group density were either negatively biased or not biased, resulting in estimates of whale density that were either not biased or slightly negatively biased.

Comparison between OK09 and SPL09 results

Examination of differences between the results from the OK09 and SPL09 methods when analysing the same iteration within a scenario led to the following observations (Figs. 7–9):

- When estimating the density of groups and whales, almost all estimates from OK09 were greater than SPL09 estimates. The relative percent difference between the estimates of whale density was on average about 10–15%, ranging from 0–25%, irrespective of their levels of bias relative to the true values. The scenarios with the largest differences included complex spatial patterns and non-synchronised diving patterns.
- There was a surprising amount of variation between the estimates from the two methods when both methods were analysing the same data. This was particularly the case for the complex scenarios

Table 3
 For the 2004 set of scenarios, true values and mean percent relative bias of the density of whales/km², density of groups/km² and mean group size, along with summary statistics for the set of scenarios (mean, median and root mean square (RMS) of the mean percent relative biases).

Scenario	True values			Mean percent relative bias											
	Density Whales	Density Groups	Mean Group Size	Density of whales				Density of groups				Mean group size			
				OK09	SPL09	Standard	OK09	SPL09	Standard	OK09	SPL09	Standard	OK09	SPL09	Standard
01	0.0715	0.0292	2.45	2.51	-9.22	-15.23	-0.71	-0.71	-7.81	-18.74	3.24	3.24	-1.47	4.29	
02	0.0706	0.0288	2.45	4.73	-20.34	-35.43	-0.09	-0.09	-11.17	-33.51	4.91	4.91	-10.08	-2.88	
03	0.0556	0.0291	1.91	11.64	-11.78	-12.78	1.49	1.49	-12.56	-13.77	10.07	10.07	0.91	1.22	
04	0.0636	0.0312	2.04	-6.81	-17.73	-27.59	-9.65	-9.65	-12.92	-29.39	3.15	3.15	-5.51	2.53	
05	0.0712	0.0291	2.45	7.79	-20.24	-27.07	5.58	5.58	-14.58	-25.44	2.15	2.15	-6.57	-2.15	
06	0.0704	0.0288	2.44	12.40	-9.21	-4.67	3.36	3.36	-10.61	-5.31	8.77	8.77	1.59	0.74	
07	0.0554	0.0291	1.91	2.89	-21.09	-43.66	-1.28	-1.28	-12.76	-42.63	4.28	4.28	-9.35	-1.66	
08	0.0633	0.0310	2.04	-2.62	-14.16	-17.15	-8.44	-8.44	-10.88	-23.44	6.47	6.47	-3.50	8.22	
09	0.0581	0.0288	2.02	9.56	-21.27	-25.60	2.36	2.36	-16.35	-23.93	7.08	7.08	-5.85	-2.14	
10	0.0571	0.0291	1.96	10.02	-8.00	1.54	3.24	3.24	-8.33	-7.11	6.62	6.62	0.37	9.36	
11	0.0727	0.0298	2.44	-17.08	-21.57	-52.16	-16.81	-16.81	-11.59	-50.73	-0.38	-0.38	-11.13	-2.92	
12	0.0707	0.0289	2.44	7.64	-11.11	-10.59	-0.55	-0.55	-9.45	-12.11	8.15	8.15	-1.81	1.68	
13	0.0584	0.0288	2.03	5.84	-12.13	-18.60	-1.99	-1.99	-8.83	-18.90	8.10	8.10	-3.43	0.45	
14	0.0566	0.0291	1.95	10.30	-21.49	-31.42	4.80	4.80	-12.76	-33.92	5.22	5.22	-9.85	3.80	
15	0.0773	0.0316	2.45	-4.41	-9.92	-19.77	-6.91	-6.91	-9.91	-22.82	2.76	2.76	0.07	4.00	
16	0.0713	0.0291	2.45	5.34	-15.60	-20.37	-0.40	-0.40	-11.88	-18.98	5.77	5.77	-4.20	-1.68	
Mean 01-16	0.0652	0.0295	2.21	3.73	-15.30	-22.53	-1.63	-1.63	-11.40	-23.80	5.40	5.40	-4.36	1.43	
Median 01-16	0.0670	0.0291	2.24	5.59	-14.88	-20.07	-0.47	-0.47	-11.38	-23.13	5.50	5.50	-3.85	0.98	
RMS 01-16	0.0656	0.0295	2.23	8.55	16.10	26.22	6.05	6.05	11.61	26.58	6.03	6.03	5.96	3.93	

Table 4
 For the 2005 set of scenarios, true values and mean percent relative bias of the density of whales/km², density of groups/km² and mean group size, along with summary statistics for the set of scenarios (mean, median and root mean square (RMS) of the mean percent relative biases).

Scenario	True values			Mean percent relative bias											
	Density Whales	Density Groups	Mean Group Size	Density of whales			Density of groups			Mean group size					
				OK09	SPL09	Standard	OK09	SPL09	Standard	OK09	SPL09	Standard			
17	0.0707	0.0288	2.45	4.21	-9.57	-19.07	-1.23	-7.91	-21.48	5.61	-1.62	3.13			
18	0.0704	0.0288	2.45	11.10	-7.08	-9.08	3.51	-9.32	-3.05	7.36	2.51	-6.27			
19	0.0742	0.0303	2.45	1.50	-6.15	-17.11	-0.95	-6.89	-20.32	2.58	0.91	4.12			
20	0.0772	0.0315	2.45	11.43	-5.52	0.71	6.14	-6.53	-3.01	5.00	1.11	3.90			
21	0.0715	0.0291	2.45	11.88	-3.37	-11.85	9.82	-10.99	-12.42	2.17	8.83	0.67			
22	0.0708	0.0289	2.45	6.09	-6.45	-17.48	3.28	-11.72	-28.55	2.87	6.26	15.55			
23	0.0713	0.0292	2.45	10.65	-5.33	0.92	9.62	-10.86	-13.98	1.15	6.37	17.46			
24	0.0738	0.0301	2.45	5.17	1.60	-10.41	3.86	-6.84	-28.11	1.48	9.43	24.69			
25	0.0770	0.0314	2.45	12.65	-3.17	2.67	12.51	-10.23	-11.13	0.33	8.09	15.64			
26	0.0739	0.0302	2.45	3.89	-3.69	-11.25	2.40	-9.76	-28.65	1.74	7.19	24.53			
27	0.0705	0.0288	2.45	15.53	-6.56	-4.00	10.96	-12.23	-11.33	4.29	6.60	8.47			
28	0.0704	0.0288	2.45	7.10	1.34	-18.90	6.31	-5.94	-28.11	1.02	8.04	12.74			
29	0.0739	0.0302	2.45	3.96	-7.72	-13.92	1.05	-6.81	-18.49	2.91	-0.86	5.61			
30	0.0711	0.0291	2.44	7.62	-4.35	-2.89	3.65	-6.98	-5.16	3.91	2.85	2.52			
31	0.0706	0.0288	2.45	2.47	-7.51	-21.91	-1.13	-8.57	-21.42	3.67	1.35	-0.65			
32	0.0714	0.0292	2.45	11.42	-6.24	-4.97	2.09	-6.96	-5.89	9.20	0.84	1.03			
Average 17-32	0.0724	0.0296	2.45	7.92	-4.99	-9.91	4.49	-8.66	-16.32	3.46	4.24	8.32			
Median 17-32	0.0714	0.0291	2.45	7.36	-5.83	-10.83	3.58	-8.24	-16.23	2.89	4.56	4.87			
RMS 17-32	0.0724	0.0296	2.45	8.91	5.79	12.52	6.17	8.89	18.66	4.17	5.56	12.14			

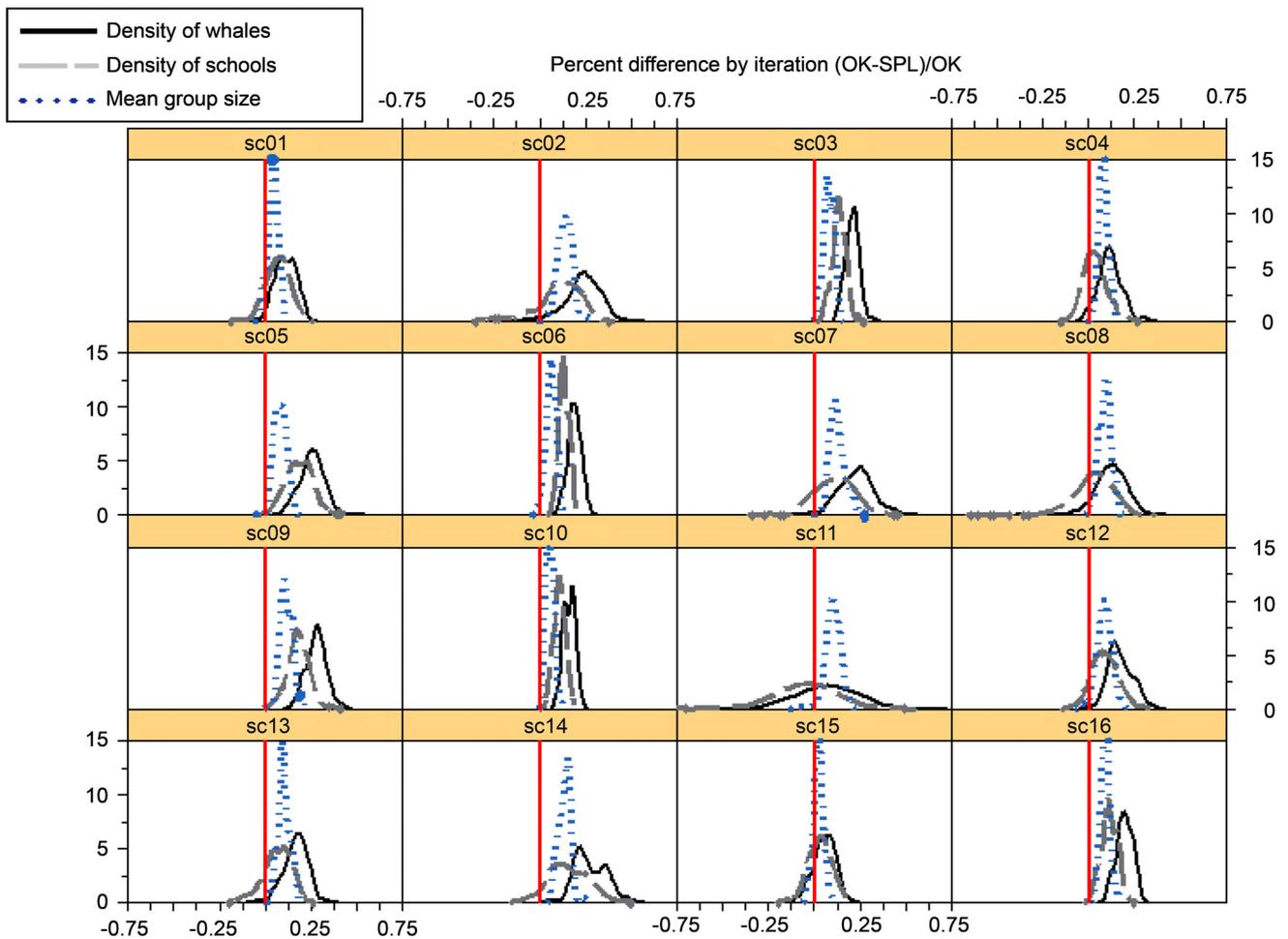


Figure 7. Distribution of the relative differences for each iteration within a scenario when comparing the results for OK09 and SPL09 methods.

that had a more realistic combination of factors, i.e., scenarios with complicated spatial correlations (sc43-50, where whale density followed a non-linear vertical (away-from-ice) gradient, group sizes followed a horizontal (along-ice) gradient, weather changed within a day and the detection probability was influenced by group size and weather conditions).

Biasing factors

Examination of the results from the ANOVAs indicated which factors were associated with the most biased estimates. Factors that were statistically significant in an ANOVA were the factors where an analysis method was deemed to be least robust, i.e., more biased density or group size estimates were commonly associated with the significant factors. The ANOVA results provided the following observations (Tables 8–10):

IM method

The IM analytical method only analysed the 2004 and 2005 scenarios (Table 8). For these sets, the IM method was the most robust and produced the least biased estimates of whale density. The main problem was dealing with the mode-changing scheme, though the extent of this problem was less than the OK method for the same scenarios.

OK method

For many scenarios, the OK09 method was able to account for the level of biases in the simulations when estimating whale density, group density and group size. The factors that were associated with the most biased estimates were as follows:

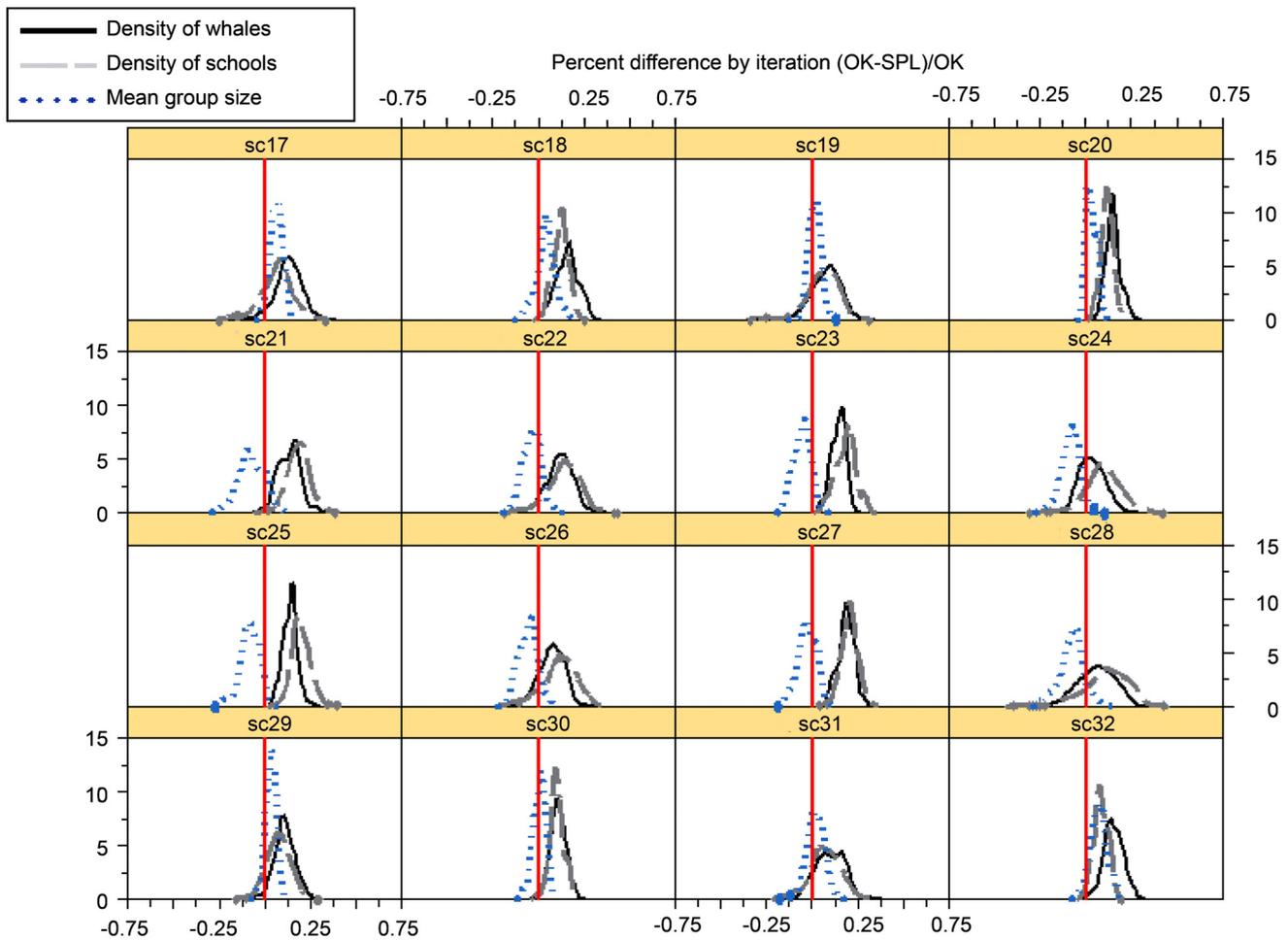


Figure 8. Distribution of the relative differences for each iteration within a scenario when comparing the results for OK09 and SPL09 methods.

- **SPATIAL GRADIENTS.** The OK09 method resulted in a biased estimate of group density and whale density when there was an interaction between weather conditions and a whale density/group size gradient (also called location correlation). The extent of bias was less than the standard estimation method but more than the SPL method (where the level of bias resulted in a non-significant factorial effect in the ANOVA).
- **NON-SYNCHRONISED DIVING.** The OK09 method resulted in biased estimates of group density, whale density and group size when the whales within a group were surfacing in a non-synchronised manner. The level of bias also depended on whether measurement error was included in the scenario. In the scenarios with measurement error and whales diving in a non-synchronised manner, the OK09 density estimates were even more positively biased than when there was no measurement error and the whales within a group dove at the same time.
- **IO/CLOSING MODE.** Alternating between IO and Closing mode was associated with biased estimates of group sizes. The level of bias of group size estimates was greater when there were group size recording errors.
- **DUPLICATES.** Misidentifying duplicate sightings led to very large positively biased estimates of whale density, group density and group size. Since there were fewer recorded duplicates than was actually the case, and the OK09 method did not attempt to account for this bias, these results are expected.

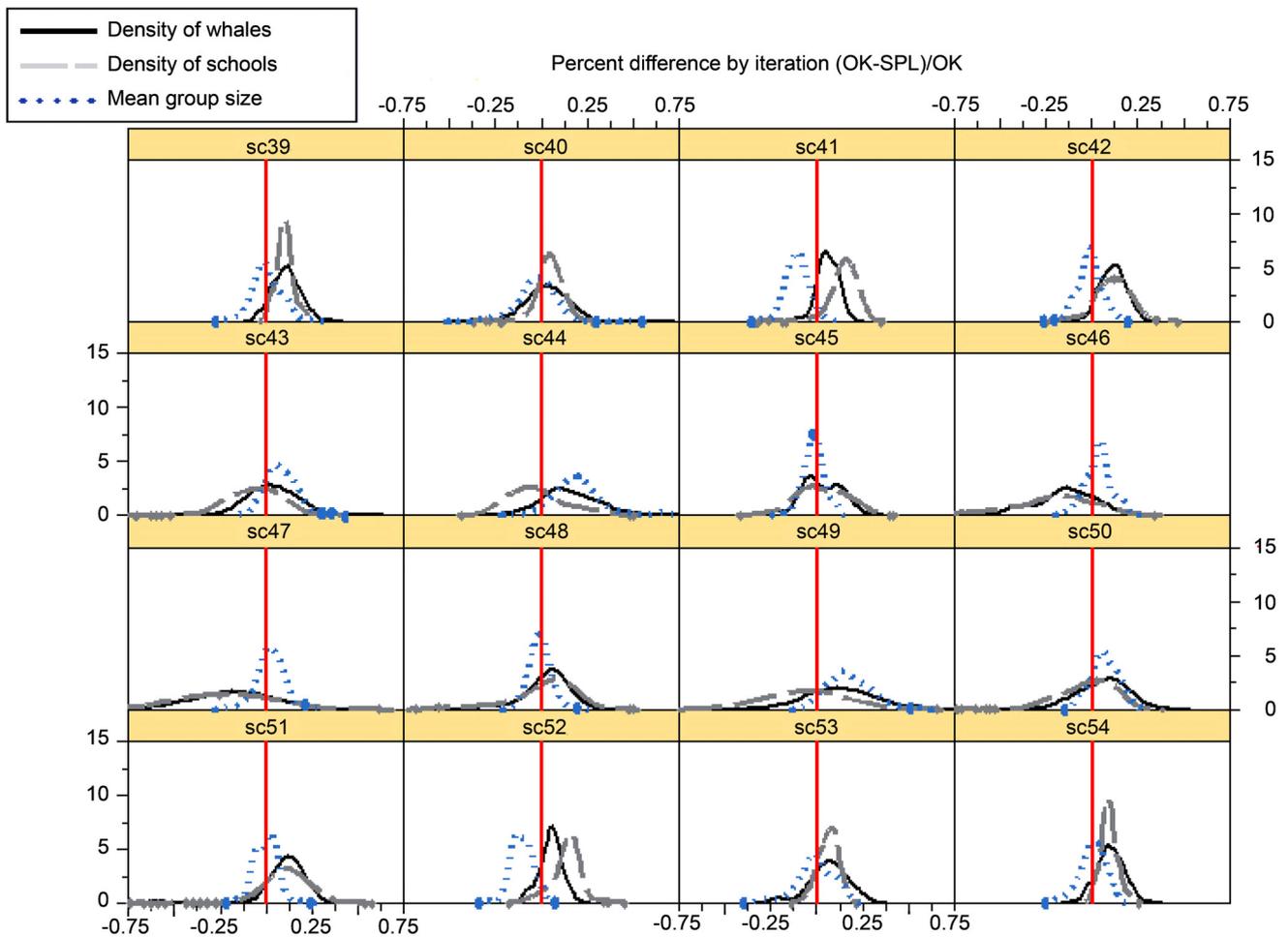


Figure 9. Distribution of the relative differences for each iteration within a scenario when comparing the results for OK09 and SPL09 methods.

SPL method

For many scenarios, the SPL09 method was able to account for the extent of biases in the simulations. The factors associated with the most biased estimates were as follows:

- **NON-SYNCHRONISED DIVING.** While there was a significant bias in the estimated whale density and group size when the whales within a group were surfacing in a non-synchronised manner in the 2005 scenarios, this bias was not significant for the 2008 scenarios. The estimates in the 2008 scenarios appear to have been compensated by the measurement errors.
- **MEASUREMENT ERRORS.** There was a significant positive bias in the estimated density of groups and whales (but not group size) when there were measurement errors in the recorded radial distance, angle, time and group size. However, if there was only measurement error in the group size, then only the group size estimate was positively biased.
- **COVARIATES IN DETECTION FUNCTION.** There was a bias in the estimates of whale density, group density and group size when the detection function depended on group size, weather or cue in the 2004 scenarios. However, the bias was less in the 2005 and 2008 scenarios when the detection function depended on group size and weather.
- **DUPLICATES.** Misidentifying duplicate sightings led to very large positively biased estimates of whale density, group density and group sizes. Since there were fewer recorded duplicates than was truly the case, and the SPL09 method did not attempt to account for this bias, these results are expected.

Table 5
 For the 2006–2007 set of scenarios, true values and mean percent relative bias of the density of whales/km², density of groups/km² and mean group size, along with summary statistics for the set of scenarios (mean, median and root mean square (RMS) of the mean percent relative biases). NA indicates not available.

Scenario	True values			Mean percent relative bias											
	Density Whales	Density Groups	Mean Group Size	Density of whales				Density of groups				Mean group size			
				OK09	SPL09	Standard	OK09	SPL09	Standard	OK09	SPL09	Standard	OK09	SPL09	Standard
33	0.0356	0.0145	2.45	4.90	-2.80	-18.40	2.88	-5.86	NA	9.20	3.36	NA	NA		
34	0.0295	0.0145	2.03	-0.08	-0.85	-30.00	-4.34	-10.08	NA	2.03	10.85	NA	NA		
35	0.1066	0.0436	2.45	7.81	-0.98	-23.04	3.69	-6.40	NA	4.83	5.81	NA	NA		
36	0.0921	0.0454	2.03	2.67	-2.40	-31.62	-4.02	-13.30	NA	4.02	13.06	NA	NA		
37	0.0361	0.0146	2.47	-1.03	-15.39	-8.99	-4.71	-15.47	NA	7.15	0.13	3.04	3.04		
38	0.0536	0.0219	2.44	7.37	-2.60	-0.10	2.26	-8.96	NA	3.91	7.05	4.00	4.00		
Average 33–38	0.0589	0.0258	2.31	3.61	-4.17	-18.69	-0.71	-10.01	NA	5.04	6.71	3.52	3.52		
Median 33–38	0.0448	0.0183	2.44	3.78	-2.50	-20.72	-0.88	-9.52	NA	4.42	6.43	3.52	3.52		
RMS 33–38	0.0660	0.0291	2.32	4.96	6.57	21.79	3.74	10.59	NA	5.69	7.99	2.05	2.05		

Table 6
 For the 2008 set of scenarios without misidentification of duplicate sightings, true values and mean percent relative bias of the density of whales/km², density of groups/km² and mean group size, along with summary statistics for the set of scenarios (mean, median and root mean square (RMS) of the mean percent relative biases).

Scenario	True values			Mean percent relative bias																				
	Density Whales	Density Groups	Mean Group Size	Density of whales				Density of groups				Mean group size												
				OK09	SPL09	Standard	OK09	SPL09	Standard	OK09	SPL09	Standard	OK09	SPL09	Standard									
39	0.1018	0.0291	3.50	11.62	-1.38	-1.67	3.75	-7.12	-4.52	7.64	6.25	2.91	0.0291	0.0291	3.50	11.62	-1.38	-1.67	3.75	-7.12	-4.52	7.64	6.25	2.91
41	0.0715	0.0292	2.45	14.87	7.59	5.23	16.49	-0.95	-5.51	-1.16	9.05	11.63	0.0292	0.0292	2.45	14.87	7.59	5.23	16.49	-0.95	-5.51	-1.16	9.05	11.63
43	0.0762	0.0291	2.62	8.83	3.86	-23.81	-8.12	-2.41	-24.77	18.31	7.01	1.34	0.0291	0.0291	2.62	8.83	3.86	-23.81	-8.12	-2.41	-24.77	18.31	7.01	1.34
45	0.0586	0.0292	2.01	-9.40	-12.18	-25.08	-16.53	-19.93	-39.71	8.87	10.25	24.43	0.0292	0.0292	2.01	-9.40	-12.18	-25.08	-16.53	-19.93	-39.71	8.87	10.25	24.43
48	0.0587	0.0291	2.02	-6.73	-9.18	-24.13	-13.51	-15.73	-38.43	8.21	8.80	23.40	0.0291	0.0291	2.02	-6.73	-9.18	-24.13	-13.51	-15.73	-38.43	8.21	8.80	23.40
50	0.0764	0.0291	2.62	9.98	4.90	-21.94	-5.94	-3.07	-23.11	16.90	8.66	1.71	0.0291	0.0291	2.62	9.98	4.90	-21.94	-5.94	-3.07	-23.11	16.90	8.66	1.71
52	0.0714	0.0291	2.45	12.98	6.93	5.16	14.66	-1.32	-6.05	-1.19	8.86	12.17	0.0291	0.0291	2.45	12.98	6.93	5.16	14.66	-1.32	-6.05	-1.19	8.86	12.17
54	0.1024	0.0291	3.52	11.98	0.11	-0.93	3.54	-6.37	-3.06	8.12	6.95	1.87	0.0291	0.0291	3.52	11.98	0.11	-0.93	3.54	-6.37	-3.06	8.12	6.95	1.87
Average 39-54*	0.0771	0.0291	2.65	6.77	0.08	-10.90	-0.71	-7.11	-18.15	8.21	8.23	9.93	0.0291	0.0291	2.65	6.77	0.08	-10.90	-0.71	-7.11	-18.15	8.21	8.23	9.93
Median 39-54*	0.0738	0.0291	2.53	10.80	1.99	-11.80	-1.20	-4.72	-14.58	8.16	8.73	7.27	0.0291	0.0291	2.53	10.80	1.99	-11.80	-1.20	-4.72	-14.58	8.16	8.73	7.27
RMS 39-54*	0.0787	0.0291	2.70	11.06	6.88	17.02	11.57	9.71	23.18	10.57	8.32	13.44	0.0291	0.0291	2.70	11.06	6.88	17.02	11.57	9.71	23.18	10.57	8.32	13.44
Average 1-54*	0.0690	0.0290	2.38	5.70	-7.59	-15.62	0.78	-9.52	-19.67	5.18	2.26	5.77	0.0290	0.0290	2.38	5.70	-7.59	-15.62	0.78	-9.52	-19.67	5.18	2.26	5.77
Median 1-54*	0.0709	0.0291	2.45	7.24	-6.82	-17.13	2.17	-9.60	-19.65	4.56	2.05	3.09	0.0291	0.0291	2.45	7.24	-6.82	-17.13	2.17	-9.60	-19.65	4.56	2.05	3.09
RMS 1-54*	0.0705	0.0294	2.40	8.80	10.76	20.15	7.15	10.27	21.45	6.51	6.59	9.41	0.0294	0.0294	2.40	8.80	10.76	20.15	7.15	10.27	21.45	6.51	6.59	9.41

*Only scenarios without misidentification of duplicates

Table 7
 For the 2008 set of scenarios with misidentification of duplicate sightings, true values and mean percent relative bias of the density of whales/km², density of groups/km² and mean group size, along with summary statistics for the set of scenarios (mean, median and root mean square (RMS) of the mean percent relative biases).

Scenario	True values			Mean percent relative bias											
				Density of whales				Density of groups				Mean group size			
	Density Whales	Density Groups	Mean Group Size	OK09	SPL09	Standard	OK09	SPL09	Standard	OK09	SPL09	Standard	OK09	SPL09	Standard
sc40	0.1022	0.0291	3.51	74.80	68.54	41.59	55.98	48.56	37.45	12.23	13.56	2.50	12.23	13.56	2.50
sc42	0.0715	0.0291	2.45	66.92	49.07	40.44	48.97	32.43	12.01	12.42	12.85	25.53	12.42	12.85	25.53
sc44	0.0766	0.0291	2.63	76.38	36.66	-3.48	23.41	26.33	-9.32	39.69	8.66	6.30	39.69	8.66	6.30
sc46	0.0587	0.0291	2.02	40.25	61.91	0.74	19.52	45.53	-18.58	17.52	12.69	23.90	17.52	12.69	23.90
sc47	0.0584	0.0291	2.01	36.29	63.65	0.61	17.18	47.03	-20.10	16.79	13.06	25.89	16.79	13.06	25.89
sc49	0.0764	0.0291	2.62	65.40	43.81	-7.55	20.54	31.29	-11.55	36.04	10.24	4.49	36.04	10.24	4.49
sc51	0.0716	0.0292	2.45	66.16	47.87	46.46	45.65	29.74	12.83	14.55	14.08	29.76	14.55	14.08	29.76
sc53	0.1022	0.0291	3.51	74.67	62.17	42.25	54.95	44.92	39.19	12.91	12.03	1.55	12.91	12.03	1.55
Mean 40-53*	0.0772	0.0291	2.65	62.61	54.21	20.13	35.78	38.23	5.24	20.27	12.15	14.99	20.27	12.15	14.99
RMS 40-53*	0.0788	0.0291	2.71	64.30	55.24	30.37	39.18	39.16	22.96	22.76	12.26	18.87	22.76	12.26	18.87

* Only scenarios with misidentification of duplicated

Table 8a

Factorial effects (difference between treatment levels) of each factor when estimating density of whales for each analysis method, in addition, the root mean square error and mean of all factorial effects

2004 (sc01-16)									
Analysis method	Density gradient	Size gradient	Clustered	Mode change	DF ¹ with size effect	DF with weather effect	DF with cue effect	Root mean square error	Mean
OK – 2004	-6.2	-1.4	2.1	-2.6	-3.3	4.9	1.9	3.59	-0.66
OK – 2007	-6.3	1.6	-0.02	-7.3	-5.1	3.8	1.0	4.43	-1.76
OK – 2008	-6.7	2.7	2.2	-6.0	-4.9	0.1	0.4	4.09	-1.75
IM – 2007	-1.8	0.2	-0.3	-3.7	0.8	-0.5	2.0	1.77	-0.45
OK – 2009	-7.1	2.7	0.2	-7.5	3.3	-3.9	3.7	4.69	-1.23
SPL – 2009	-1.0	-1.3	-1.1	0.02	9.2*	-2.2*	-2.0*	3.73	0.23
Standard	-5.2	1.3	0.2	-2.1	20.8*	-11.0*	0.1	9.16	0.60

Table 8b

2005 (sc17-32)							
Analysis method	Group size error	Weather and density interaction	Vertical and horizontal density	Mode changes	Non-synchronised surfacing	Root mean square error	Mean
OK – 2005	-1.3	14.3*+	1.2	-10.3+	1.5	7.95	1.08
OK – 2007	-0.1	5.2*	0.7	1.1	0.2	2.40	1.42
OK – 2008	-0.9	-3.7*	1.4	2.5*	2.6*	2.43	0.38
IM – 2007	-0.4+	4.0	1.8	-4.8+	-0.5	2.92	0.02
OK – 2009	-1.0	-7.2*	1.3	1.6	2.4*	3.55	-0.58
SPL – 2009	2.8	0.4	-0.6	-1.4	3.6*	2.16	0.96
Standard	-2.6	-12.7*	1.4	-7.0*	2.2	6.70	-3.74

Table 8c

2008 (sc39-54)							
Analysis Method	Measure errors	Along ice density gradient	Mis-id dups	Non-synchronised surfacing	Spatial correlation	Root mean square error**	Mean**
OK – 2008	2.7+	1.3	53.7*+	-10.1*	-11.2*+	7.69	-4.33
OK – 2009	-1.2+	2.4	55.8*+	-14.0*	-14.1*	10.03	-6.73
SPL – 2009	15.6*+	-1.7	54.1*+	-0.4	-5.9	8.38	1.90
Standard	3.2+	1.6	31.0*	3.1+	-35.4*	17.86	-6.88

¹ DF = detection function

* Level of bias was significant at the 0.01 p-value level

** Mis-identification of duplicates factor not included

+ Significant interaction between the two factors within the same row

DISCUSSION

Comparisons between the estimated and true density values highlighted several factors associated with biased estimates resulting from the new analysis methods when applied to the simulated data. These results helped developers create improved methods that, when applied to the actual data, led to the eventual adoption of abundance estimates approved by the Scientific Committee which use a combination of the OK and SPL methods (IWC, 2013a).

When analysing the simulated data, the factors associated with the most biased OK09 estimates of whale density were spatial gradients and non-synchronised diving patterns. The factors associated with the most biased SPL09 estimates were measurement errors and non-synchronised diving patterns. These same factors turned out to be important in understanding the differences between the abundance estimates resulting from the OK and SPL methods when applied to the actual data.

After many years spent investigating the intricacies of the OK and SPL methods and the available actual data, the Scientific Committee agreed to adjust downward the results for the OK method due to the inability of the OK model to accurately estimate several components. These components were: (1) the inter-cue interval (dive

Table 9a

Factorial effects (difference between treatment levels) of each factor when estimating the mean group size for each analysis method, in addition, the root mean square error and mean of all factorial effects

2004 (sc01-16)									
Analysis method	Density gradient	Size gradient	Clustered	Mode change	DF with size effect	DF with weather effect	DF with cue effect	Root mean square error	Mean
OK – 2009	-0.6	2.0	-0.01	-3.5*	2.7*	-0.8	0.9	1.90	0.10
SPL – 2009	-0.7	-0.3	-0.3	-0.7	6.9*	-3.9*	-0.4	3.03	0.09

Table 9b

2005 (sc17-32)							
Analysis method	Group size error	Weather and density interaction	Vertical and horizontal density	Mode changes	Non-synchronised surfacing	Root mean square error	Mean
OK – 2009	-1.3+	-1.4+	0.1+	2.1*+	-3.1*+	1.88	-0.72
SPL – 2009	2.0*	-0.8	0.1	-0.3	6.7*	3.15	1.54

Table 9c

2008 (sc39-54)							
Analysis method	Measure errors	Along ice density gradient	Mis-id dups	Surfacing patterns	Location correlations	Root mean square error**	Mean**
OK – 2009	-5.4+	0.6	12.1*+	-9.5*	12.1*	9.08	1.98
SPL – 2009	0.9	-0.3	3.9*	2.0*	-0.5	2.02	1.20

¹ DF = Detection function
 * Level of bias was significant at the 0.01 p-value level
 ** Misidentification of duplicates factor not included
 + Significant interaction between two factors within the same row

Table 10a

Factorial effects (difference between treatment levels) of each factor when estimating the density of groups of whales for each analysis method, in addition, the root mean square error and mean of all factorial effects

2004 (sc01-16)									
Analysis method	Density gradient	Size gradient	Clustered	Mode change	DF with size effect	DF with weather effect	DF with cue effect	Root mean square error	Mean
OK – 2009	-6.3	0.9	0.1	-4.0	0.6	-3.0	2.8	3.25	-1.27
SPL – 2009	-0.3	-1.1	-0.8	0.6	3.2*	1.5	-1.6	1.57	0.21

Table 10b

2005 (sc17-32)							
Analysis method	Group size error	Weather and density interaction	Vertical and horizontal density	Mode changes	Non-synchronised surfacing	Root mean square error	Mean
OK – 2009	0.4	-5.6*	1.1	-0.6	5.7*	3.62	0.20
SPL – 2009	0.9	1.2	-0.7	-1.1	-2.3	1.36	-0.40

Table 10c

2008 (sc39-54)							
Analysis method	Measure errors	Along ice density gradient	Mis-id dups	Surfacing patterns	Location correlation	Root mean square error**	Mean**
OK – 2009	6.1	1.0	36.5*	-2.0	-25.9*	13.35	-5.20
SPL – 2009	13.5*	-1.2	45.3*	-1.9	-3.9	7.12	1.63

¹ DF = detection function
 * Level of bias was significant at the 0.01 p-value level
 ** Misidentification of duplicates factor not included
 + Significant interaction between two factors within the same row

time) from the available survey data that contain rounded values of the initial detection time; (2) the true sizes of groups from the group sizes recorded in IO and Closing modes; and (3) the intra-stratum scale density spatial patterns (IWC, 2013a). In order to address these issues, the Scientific Committee agreed on methodological changes to the OK method. It was agreed to account for Component 1 by using fixed values for the inter-cue intervals (dive times) derived from field experiments (Hedley, 2012). It was agreed to account for Component 2 by using group size error correction factors derived from field experiments (IWC, 2013b). The Scientific Committee also agreed to adjust correction factors for Components 1 (treatment of confirmed and unconfirmed group sizes derived from the SPL method) and 3 (treatment of the spatial imbalance derived from the SPL method).

When analysing the simulated data, the relative percent differences of the OK09 whale density estimates were on average about 10–15% greater than the corresponding estimates from the SPL09 methods, irrespective of their levels of bias relative to the true values. The mode of the relative percent differences of the OK09 results generally ranged from 0–25% greater than the corresponding SPL09 results.

In comparison, when analysing the actual data, the OK method resulted in greater abundance estimates as compared to the SPL method, though we will never know the level of bias for either method relative to the true abundance. In the earlier versions (2009–11) of the OK and SPL methods when applied to the actual data, the relative percent difference ranged from 30–60%, depending on the version of each method, how the standardised input data were processed and the CP data series. The differences between the two methods were always greater when analysing the CPII data compared with the CPIII data because some of the potential biasing attributes of the data collection were corrected in the CPIII time series. The resulting differences in abundance estimates were smaller by the time the final versions of the OK and SPL methods were developed. This can be seen by the smaller relative percent differences of 34% for CPII and 6% for CPIII between the abundance estimates from the OK method without the agreed methodological changes (see Table 4, Okamura & Kitakado, in prep) and the estimates with the agreed adjustments (IWC, 2013a).

In summary, the simulation exercise identified factors that were associated with biased estimates. Some of these factors were also detected when analysing the actual data. While a good place to start, the initial simulations lacked complexity as the factors included were acting independently. Later sets of scenarios were more realistic, with biasing factors occurring simultaneously and generating interactions. But the simulations were still a simplification of the intricacies of the actual data. In addition, the methods used to analyse the simulated data were simplifications of the versions used to analyse the actual data. In order to create a more realistic simulated dataset, the simulation scenario sc55 was developed to incorporate into a single scenario most of the interacting complex biasing factors that had been investigated in this simulation exercise. In the future, if the most current versions of the OK and SPL methods analyse this new scenario, we could gain further insight into the robustness of these methods and adequacy of the simulations.

REFERENCES

- Branch, T.A. (2006). Abundance estimates for Antarctic minke whales from three completed sets of circumpolar surveys. SC/58/IA18 presented to the IWC Scientific Committee, St. Kitts and Nevis, 2006. [Available from the IWC Publications Team]
- Bravington, M.V., & Hedley, S. (2010). Antarctic minke whale abundance from the SPLINTR model: Some ‘reference’ dataset results and ‘preferred’ estimates from the second and third circumpolar IDCR/SOWER surveys. SC/62/IA12 presented to the IWC Scientific Committee, Agadir, Morocco, 2010. [Available from the IWC Publications Team]
- Cooke, J.G. (2002). Progress report on analysis of IDCR/SOWER data. SC/54/IA1 presented to the IWC Scientific Committee, Shimonoseki, Japan, 2002. [Available from the IWC Publications Team]
- Hedley, S.L. (2012). Respiration rates of Antarctic minke whales from the Visual Dive Time Experiment on the 2004/05 IWC SOWER Survey. SC/64/IA12 presented to the IWC Scientific Committee, Panama City, Panama, 2012. [Available from the IWC Publications Team]
- International Whaling Commission (2013a). Report of the Scientific Committee: Annex G. *J. Cetacean Res. Manage.* Suppl. 13: 175–91.
- International Whaling Commission (2013b). Report of the 2012 Intersessional IA workshop on estimating abundance of Antarctic minke whales: Annex E. *J. Cetacean Res. Manage.* Suppl. 13: 361–68.
- NIST/SEMATECH e-Handbook of Statistical Methods. [Available at: <https://doi.org/10.18434/M32189>]
- Okamura, H., & Kitakado, T. (2010). Abundance estimates of Antarctic minke whales from the historical IDCR/SOWER survey data using the OK method. SC/62/IA3 presented to the IWC Scientific Committee, Agadir, Morocco, 2010. [Available from the IWC Publications Team]
- Okamura, H., & Kitakado, T. (in prep). A line-transect method accounting for uncertain detection on the trackline and abundance of Antarctic minke whales from the second and third circumpolar IDCR/SOWER surveys. *J. Cetacean Res. Manage.* Special Issue 4.
- Petersen, R.G. (1985). *Design and Analysis of Experiments*. Marcel Dekker, New York.

- Polacheck, T., Gray, R., & Klaer, N. (1996). Documentation for TRANSIM: A line-transect survey simulation model for whales. SC/A96/AE18 presented to the IWC Scientific Committee, Aberdeen, UK, 1996. [Available from the IWC Publications Team]
- Polacheck, T., Palka, D., Borchers, D., Cooke, J.G., Skaug, H.J., & Dixon, C. (2000). Simulation comparison of different methods for estimating whale density from double-team line transect data. SC/52/RMP18 presented to the IWC Scientific Committee, Adelaide, Australia, 2000. [Available from the IWC Publications Team]
- Skaug, H.J., & Schweder, T. (1999). Hazard models for line transect surveys with independent observers. *Biometrics* 55(1): 29–36. [Available at: <https://doi.org/10.1111/j.0006-341x.1999.00029.x>]
- Ward, A.J. (1988). Accounting for the effect of pod size in the estimation of blow rates: Analysis of data from the 1984/85, 85/86 and 86/87 IDCR minke whale assessment cruises. *Rep. Int. Whal. Comm.* 38: 265–8.

©Authors. This is an open access article distributed under the terms of a *Creative Commons License CC-BY-NC 4.0*.

Supplementary Material

DESCRIPTION OF SIMULATED DATA

The software comprises seven components which deal with the following issues:

1. Distribute whales in space (pre-survey);
2. Set rotation schedules for observer teams within a platform and set periods when specific detection functions are associated with a team (pre-survey);
3. Control location of the ship;
4. Control movement and diving/surfacing behaviours of the whales;
5. Determine whether a whale was detected by a specified platform at a given surfacing;
6. Write output files;
7. Control post-processing issues: introduce measurement error into the radial sighting distance, sighting angle and time of detection data; remove from the final output file those sightings detected by the Bridge platform after the animal group has been seen by one or both of the other platforms; identify whales detected during Closing mode.

1. Whale distribution

In general, the survey area was set up within an x–y grid pattern, where the ice edge was at $y = 0$ and the survey track lines were perpendicular to the ice edge following a constant x value (Figs. 1 and 3). Thus, x was the distance along the ice edge and y was the distance away from the ice edge.

The underlying distribution of whales was set up according to either: (a) a uniform random distribution; (b) a Neyman Scott clustering process; (c) gradients in either group size or density; or (d) a combination of these. In addition, there could be a correlation between weather conditions and density of whales.

The Neyman Scott clustering process characterised individual whale groups within clusters of whale groups by their distance from cluster ‘parent’ points, where the points were not themselves whale group locations but approximately the ‘centre’ of these clusters which could either be distributed uniformly or according to a density gradient. The numbers of ‘parent’ points were determined by the ratio between density of whale groups and density of clusters, both of which were user-specified. The distribution of whale group distances from cluster ‘parent’ points was determined by the ‘standard deviation within clusters’ which is also user-specified.

Group size gradients were produced as gradients between user-specified group size distributions for the start and end of the whale distribution area. In such cases, the group sizes near the ice edge were large, while group sizes farthest from the ice edge were small (Fig. 1).

Density gradients were created according to an equation that employed a user-specified density gradient parameter to ‘redistribute’ whales from an initial uniform distribution to a distribution that tended to bunch them at one end of the survey area. In such cases, the density was highest near the ice edge and lowest far from the ice edge (Fig. 3). The overall density within the entire area remained as specified by the user in ‘density of whales’. The resulting density gradient could be linear, monotonic or non-monotonic. For example, a monotonic gradient in density of whales in the direction away from the ice edge (parallel to the y -axis, e.g., Figs. 1 and 3) was generated through the equation: $y' = 1.0 - (y^m)$, where y was the initial position of the group on the y -axis and y' was the modified position on the y -axis. For $m = 0$, there was no redistribution and the whale distribution remained uniform. For $0 < m < 1$, the result was a redistribution towards a higher density in the near-to-ice area while preserving the overall number of whales. The larger the value of m , the farther y' is from the original position y (Fig. S1).

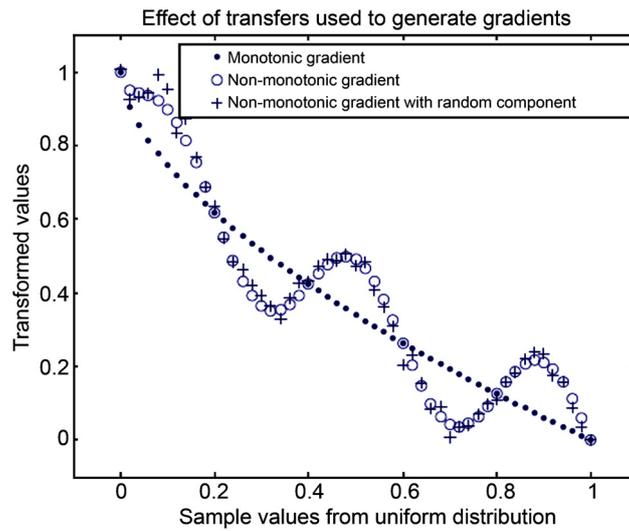


Figure S1. Effects of three transformations used to generate different types of group density spatial gradients.

In order to achieve a non-monotonic density gradient, a sinusoidal component was added to the transform (Fig. S1). A small random amount taken from a normal distribution with a mean of zero and standard deviation of 0.02 was added as follows to smooth this gradient:

$$y' = 1.0 - (y^m) + [\alpha \cdot \sin(5y \pi)] + \text{Normal}(0, 0.02^2)$$

In practice, the program code constrained the choice of α given a particular value of m to exclude negative values of y' . In order to generate the position of a whale group, the coordinate within the unit interval resulting from one of these transformations was multiplied by the actual y dimension of the whale distribution area. In these simulations, $m = 0.6$ for creation of density gradients and $\alpha = 0.15$ for creation of non-monotonic gradients.

All the possible distribution patterns above were used in one or more scenarios. In some scenarios, the gradient in underlying density of whales was in both the away-from-ice (vertical y) and along-ice (horizontal x) directions, where the away-from-ice gradient could be non-linear (Fig. 1). In cases with gradients in both directions, the whale density was highest in the lower (near the ice) left-hand corner of the study area, lowest in the upper (far from ice) right-hand corner of the study area, and intermediate at both the upper (far from ice) left-hand corner and lower (near the ice) right-hand corner. However, the densities at the two corners with intermediate densities were not necessarily the same.

There were also scenarios which involved a correlation between weather conditions and the density of whales or whale groups (Fig. 3). In these scenarios, it was assumed that in areas where weather conditions were worst (far from the ice edge), the true whale density was lowest, and in areas where weather conditions were best (at the ice edge), the true whale density was the highest, which is in line with the actual experience on the cruises (Murase *et al.*, 2020). In the simulations, the weather pattern was as follows for each day: surveying started at the ice edge and progressed northward away the ice edge, encountering 1/6th best weather conditions, 1/2nd intermediate weather conditions and 1/3rd worst weather conditions, in this order, over the total daily survey time. Thus, over a 16-hour survey day, for the first ca. two hours and 40 minutes, surveyors experienced the best weather conditions (code = 0); for the next c. eight hours, they had intermediate weather conditions (code = 1); and for the last c. five hours and 20 minutes, they had the worst weather conditions (code = 2). Note that these times of day were approximate in scenarios including Closing mode because diversions to close on groups could delay the time at which the ship returned to the track line.

2. Survey protocol

The software allowed input of the following user-specified logistic protocols: number of ships involved in the survey; number of platforms per ship; number of independent teams per platform (where a team is a set of

observers that can have a team-dependent detection function); number of survey hours per day; number of days surveyed; and speed of the ship(s). In addition, the user could specify the following data collection protocols: rotation schedule of teams; furthest distance from the ship at which a sighting could be seen by a team on a platform; type of searching protocol (IO¹ vs. Closing² mode); and whether all surfacings or only the initial surfacing of a detected group was recorded in the output file.

In these simulations, logistic protocols involved collecting data from a one-ship, three-platform survey. The three observation platforms were: (a) the main platform, called the Topman team; (b) the additional team needed under the IO mode only, called the IO team; and (c) the team located on the Bridge, called the Bridge team. The Bridge team recorded all data from the three teams, searched for whales and determined whether a group was a duplicate (seen by multiple teams). There were two teams of observers surveying on each platform. Teams rotated within the platforms every two hours, on the hour. A team always stayed on the same platform. Each simulated dataset consisted of multiple survey days, in which observers searched for 16–22 hours per day while the ship travelled at 11.5 knots when on the track line. The number of survey days varied between different simulation scenarios and the total number of sightings within an individual dataset varied within a scenario. The number of sightings within these datasets was larger than typically seen in an actual survey because of the aim to test the robustness of the analysis methods without restricting sample size. In the future, datasets with different sample sizes could be analysed to investigate sample size effects.

In these simulations, data collection protocols were set so that the output file included only the initial surfacing of a detected group from each platform. Because observers used low-powered binoculars (7 × 50), it was possible for an observer to detect a group 9,000 m (4.9 n.mi) away from the track line (Fig. S2). Groups were detected from 0° (on the track line) to 90° on either side of the track line. Most sightings made by an observer were within a perpendicular distance of 1.5 n.mi, as shown in Branch and Butterworth (2001). During Closing mode, when

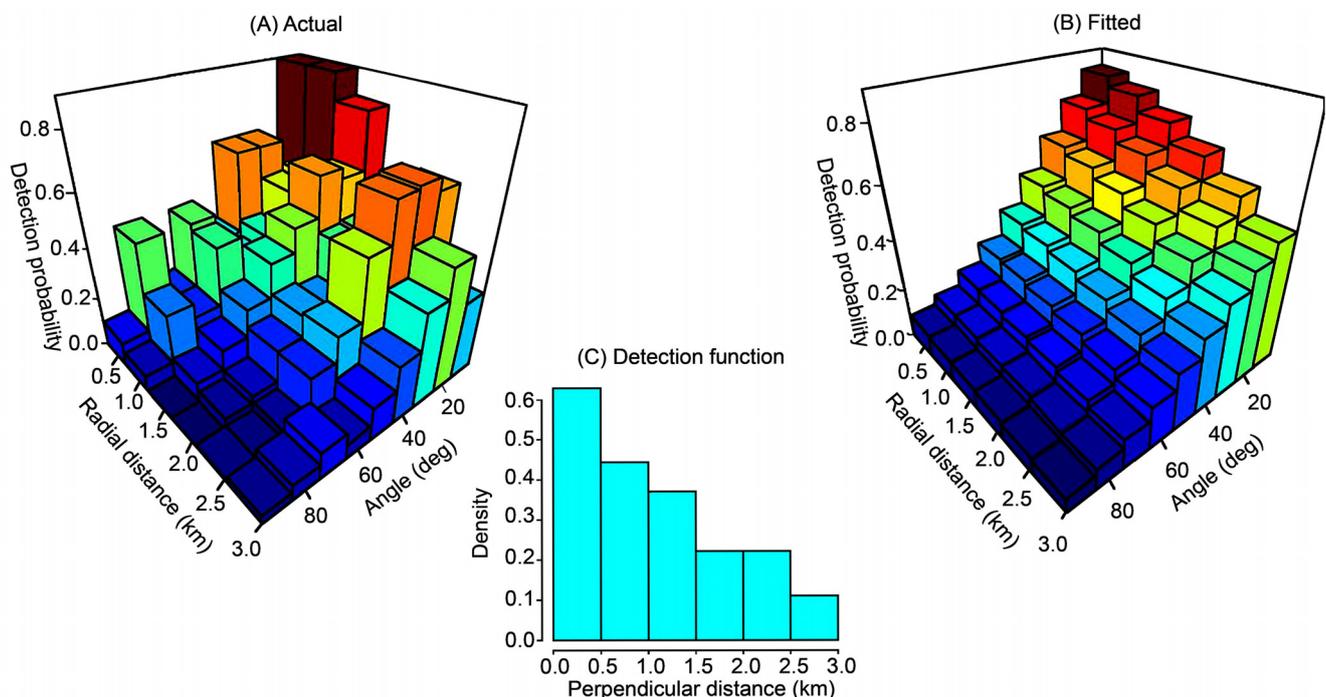


Figure S2. The detection function for the Topman team for scenario sc55. (A, B) Function described by the probability of detecting a group (vertical z-axis) given its position as measured by the sighting angle (x-axis) and radial distance (y-axis). Parameter estimates (Table S5) for the detection function were derived from a distribution of the actual recorded IWC IDCR/SOWER data as shown in (A). The resulting smoothed predicted three-dimensional detection function used in the simulation (B) and two-dimensional perpendicular distance from the track line detection function (C).

¹ IO mode (a.k.a. passing mode) occurred when both the Topman and IO teams (plus the Bridge Team) were surveying, and the ship stayed on the track line so that no whales were approached and therefore group sizes were unconfirmed.

² Closing mode occurred when only the Topman and Bridge teams were surveying and groups of whales that were detected 0.3–3.0 n.mi from the track line were approached to confirm the group size.

only the Topman and Bridge teams were in place, the time the ship spent paused near a whale group to confirm its size varied, with a mean of seven minutes and a standard deviation of one minute. Most aspects of Closing mode were implemented in the post-process component.

In some scenarios, IO survey mode was the only survey procedure used. In other scenarios, IO mode alternated every four hours with Closing mode. During Closing mode, groups detected between 0.3–3 n.mi perpendicular distance from the track line were approached and the group size confirmed. This procedure involved the observers going off-effort, the ship approaching the group and the observers then determining the correct group size for the entire group. In the 2004 simulations, for scenarios where survey modes alternated, each day started with Closing mode. After 2004, for scenarios that involved alternating IO and Closing mode, the mode at the start of the day also alternated. On odd-numbered days, the day started in IO mode and then alternated between modes approximately every four hours. On even-numbered days, the day started in Closing mode and alternated every four hours.

3. Location of ship

The software allowed for one or more ships to travel along a track line at a user-specified speed that was constant while on the track line. In order to avoid possible edge effects in these simulations, the ship started at a user-specified position on the track line away from the edge of the simulated area. It was possible to survey a series of parallel track lines and provide a relative X and Y position of the ship, similar to a latitude and longitude position.

In these simulations, the parallel track lines were set up perpendicular to an 'ice edge'. For example, surveying started on any given day near the ice edge ($Y = \text{small}$) and the ship travelled directly away from the ice edge, on a constant X coordinate, towards a large value of Y . On the next day, the ship moved over 20,000 m in the X -direction, started again at the ice edge ($Y = \text{small}$), then travelled directly away from the ice edge towards $Y = \text{large}$. In order to avoid potential edge effects based on how the study area was set up to contain whales in various locations, surveying started 10 n.mi (18,532.5 m) from the edge of the simulated area. Thus, the 'ice edge' was at $Y = 18,532.5$ for all X 's. Putting this all together, in a scenario where only IO mode was conducted, Day 1 started at location $X = 0$ $Y = 18,532.5$ and the ship travelled for 16 hours to end at $X = 0$ $Y = 359,530$. Day 2 started at $X = 20,000$ $Y = 18,532.5$ and the ship travelled for 16 hours to end at $X = 20,000$ $Y = 359,530$. Day 3 started at $X = 40,000$ $Y = 18,532.5$ and the ship travelled for 16 hours to end at $X = 40,000$ $Y = 359,530$, and so on. If Closing mode was in operation, then due to time spent off track to close on whale groups, the end point for a day might be less than 359,530 or the end time might be later than 16:00:00, but the ship always worked for at least 16 hours.

4. Whale movement and dive pattern

The software allowed whale groups to dive, swim, move randomly or react to the ship. In the absence of any reactive movement, each group could move at a constant speed or not at all. In this set of simulations, the groups did not move. The speed could be constant over all whales as in the present simulations where it was zero or it could be taken from a distribution with a user-specified variance.

Diving and surfacing times could be generated from either a Poisson distribution or an empirical list of sequences. Individuals within whale groups of any size could either dive and surface in unison (synchronised) or individuals within a group could surface at different times from other animals in the same group (non-synchronised). When an animal within a group surfaced, it remained at the surface for one second, at which time it was determined whether any of the teams detected it. If the user specified that reactive movement occurred, then once a whale group entered into a 'reactive region' as defined by a user-specified distance from the ship, the whale group reacted by changes in its dive time, swim speed and/or swim direction, resulting in an altered surfacing position for the next surfacing. The amount of reaction was a function of the user-specified reactive distance and the distance between the whale and ship. The formulas for the reaction factor and changes in dive time, swim speed and swim direction were developed by Popenoe (2000) and are shown in Appendix 1.

In these simulations, whale groups did not swim or react to the ship. In scenarios sc01-54, when diving was synchronized, the dive time interval followed a Poisson process with a mean of 75 seconds (Ward, 1988), restricted to a range of 5–480 seconds (Fig. 2).

In scenarios with a Poisson surfacing distribution and non-synchronised diving patterns, individual whales within a group of two or more might not all surface at the same time, so not all individuals were available to be detected by the observers. The non-synchronised dive patterns were implemented as follows. For groups of size greater than one, each whale within the group was randomly assigned a dive time from a ‘starting’ distribution. The animal with the minimum of these dive times was then chosen and it was assumed that at least this one whale in the group surfaced. A probability that any one of the other whales also surfaced at the same time was user-specified. In these scenarios, this probability was 0.50. The data generation process ensured that each individual whale came to the surface at least once every 480 seconds. For all group sizes in scenarios where diving was non-synchronised, the average dive time interval was 75.2 seconds.

For scenario sc55, empirical dive sequences were taken from the ‘Visual Dive Time Experiment’ conducted in Antarctic waters during the 2004/05 season (Hedley & Ensor, 2006; Hedley, 2012). For each group of a given size, an empirical sequence observed for a group of that size was chosen randomly, and a starting point within that sequence was also randomly chosen. The sequence was followed from that point, but if it was exhausted before the ship passed the whale group in question, the sequence was ‘wrapped around’ to start at the beginning again. Since the largest group size from the experiment was six, the empirical sequence for group size six was used in the simulations for all groups of six or more.

From the empirical surfacing sequences, the number of whales in a group that were defined as ‘at the surface at once’ was determined from the number seen to surface within three seconds of each other in the empirical data. A single surfacing time at approximately the midpoint of such a ‘blurred’ surfacing was assigned to the event, and any seconds ‘left over’ were split between the preceding and following intervals.

In scenarios with synchronised diving where the detection function depended on group size, and in all scenarios with non-synchronised diving, the probability of detecting a group depended on the number of individuals at the surface, which might not be the total number of animals in the group.

If there was no measurement error in the group size, then the group size recorded in the sighting file was the actual total group size, not the number of individuals at the surface at any individual surfacing. This assumed that the observers were able to assimilate all the surfacing information, determine the group size accurately and record this accumulated group size information together with the time of the initial surfacing. If there was measurement error under both synchronised and non-synchronised diving, the recorded value of the group size was the total group size with the measurement error added.

5. Detection of a whale group

The user can specify that the probability of detecting a group of whales depends on the radial distance between the ship and the group, angle between the track line and line of sight to the group, platform, team within a platform, group size, weather conditions, cue, or a combination of these factors.

In these simulations, the probability of detection depended on each of these possibilities in one or more scenarios, with the exception of team within a platform as there was only one team per platform. Cue had two levels: either body or blow (not both). There were three levels of weather conditions: Level 0 representing Beaufort sea state 0; Level 1 representing Beaufort 1 and 2; and Level 2 representing Beaufort 3 and above. The length of time surveying within a weather condition level was also user-specified. The detection function model used in these simulations was an inverse logit that was a function of radial distance, sighting angle, group size and cue:

$$Pr(u | R, \theta) = \frac{\frac{\exp(b_1(R - (b_2 + b_6s^5 + b_8c)))}{\exp(b_1(R - (b_2 + b_6s^5 + b_8c))) + 1} \frac{\exp(b_3(\theta - b_4))}{\exp(b_3(\theta - b_4)) + 1}}{\frac{\exp(b_1(b_2 + b_6s^5 + b_8c))}{\exp(b_1(b_2 + b_6s^5 + b_8c)) + 1} \frac{\exp(b_3b_4)}{\exp(b_3b_4) + 1} \frac{\exp(b_5 + b_7s^5 + b_9c)}{\exp(b_5 + b_7s^5 + b_9c) + 1}}$$

where Pr = probability of detecting a surfacing given that the whale group had not been detected before,

R = radial distance between the group and the ship (meters)

θ = sighting angle relative to the track line (radians)

b_i = user specified parameters, where $i = 1$ to 9

s = group size

c = cue, where $c = 1$ = blow and $c = 0$ = body

Parameters b_1 and b_2 affected the radial distance component; b_3 and b_4 affected the angle component; and b_5 affected the overall level. Parameter b_6 added a group size effect to the radial distance component and b_7 added a group size effect to the overall level component. Parameter b_8 added a cue effect in the radial distance component and b_9 added a cue effect in the overall level component.

As in the actual SOWER data, the distributions of initial sightings detected by the IO platform were on average slightly closer to the ship and closer to the track line than those detected by the Topman platform. The Bridge sightings were on average detected even closer to the ship and the track line than the IO sightings.

Table S1

For the 2004 set of scenarios (sc01-16), the presence, absence or description of factors incorporated in each scenario. 'Present' in the DF-Cue column is interpreted as the probability of detecting a group as a function of cue

Scenario	Factors							Detection Function Number
	Density Gradient	Size Gradient	Groups Clustered	Mode	DF-size	DF-weather	DF-Cue	
1	Absent	Absent	Random	IO + Close	Present	Present	Absent	2.0–2.2
2	Present	Absent	Random	IO	Absent	Present	Present	3.0–3.2
3	Absent	Present	Random	IO	Present	Absent	Present	4.0
4	Present	Present	Random	IO + Close	Absent	Absent	Absent	4.1
5	Absent	Absent	Clumped	IO + Close	Absent	Absent	Present	4.2
6	Present	Absent	Clumped	IO	Present	Absent	Absent	2.0+
7	Absent	Present	Clumped	IO	Absent	Present	Absent	1.0–1.2
8	Present	Present	Clumped	IO + Close	Present	Present	Present	5.0–5.2
9	Present	Present	Clumped	IO	Absent	Absent	Present	4.2
10	Absent	Present	Clumped	IO + Close	Present	Absent	Absent	2.0+
11	Present	Absent	Clumped	IO + Close	Absent	Present	Absent	1.0–1.2
12	Absent	Absent	Clumped	IO	Present	Present	Present	5.0–5.2
13	Present	Present	Random	IO	Present	Present	Absent	2.0–2.2
14	Absent	Present	Random	IO + Close	Absent	Present	Present	3.0–3.2
15	Present	Absent	Random	IO + Close	Present	Absent	Present	4.0
16	Absent	Absent	Random	IO	Absent	Absent	Absent	4.1

Table S2

For the 2005 set of scenarios (sc17-32), the presence, absence or description of factors incorporated in each scenario

Scenario	Factors					Detection Function Number
	Group Size Errors	Weather & Density	Vertical & Horizontal Density	Mode	Non-synch Diving	
17	Absent	Present	Absent	IO	Absent	2.0–2.2
18	Present	Absent	Present	IO	Absent	2.0
19	Present	Present	Absent	IO + Close	Absent	2.0–2.2
20	Absent	Absent	Present	IO + Close	Absent	2.0
21	Present	Absent	Absent	IO	Present	2.0
22	Absent	Present	Present	IO	Present	2.0–2.2
23	Absent	Absent	Absent	IO + Close	Present	2.0
24	Present	Present	Present	IO + Close	Present	2.0–2.2
25	Present	Absent	Present	IO + Close	Present	2.0
26	Absent	Present	Absent	IO + Close	Present	2.0–2.2
27	Absent	Absent	Present	IO	Present	2.0
28	Present	Present	Absent	IO	Present	2.0–2.2
29	Absent	Present	Present	IO + Close	Absent	2.0–2.2
30	Present	Absent	Absent	IO + Close	Absent	2.0
31	Present	Present	Present	IO	Absent	2.0–2.2
32	Absent	Absent	Absent	IO	Absent	2.0

Table S3

For the 2006-07 set of scenarios (sc33-38), the presence, absence or description of factors incorporated in each scenario. 'Present' in the DF-Weather column is interpreted as the probability of detecting a group as a function of the weather condition

Scenario	Factors							Detection Function Number
	Group Size Errors	DF-Weather	Density Gradient	Crossways Density Gradient	Group size Gradient	Non-synch Diving	Errors in radial and angles	
33	Present	Absent	Absent	Absent	Absent	Absent	Absent	2.0
34	Present	Present	Monotonic	Present	Present	Present	Absent	2.0–2.2
35	Present	Absent	Absent	Absent	Absent	Absent	Absent	2.0
36	Present	Present	Monotonic	Present	Present	Present	Absent	2.0–2.2
37	Absent	Absent	Non-monotonic	Absent	Absent	Absent	Absent	2.0
38	Present	Absent	Non-monotonic	Absent	Absent	Absent	Present	2.0

Table S4

For the 2008 set of scenarios (sc39-54), the presence, absence or description of factors incorporated in each scenario

Scenario	Factors					Detection Function Number
	Measure error	Along Ice Density Gradient	Mis-identifying Duplicates	Surfacing Patterns	Location correlations	
39	Absent	Present	Absent	In-synch	Absent	2.0
40	Present	Absent	Present	In-synch	Absent	2.0
41	Present	Present	Absent	Out-of-synch	Absent	2.0
42	Absent	Absent	Present	Out-of-synch	Absent	2.0
43	Present	Absent	Absent	In-synch	Present	2.0–2.2
44	Absent	Present	Present	In-synch	Present	2.0–2.2
45	Absent	Absent	Absent	Out-of-synch	Present	2.0–2.2
46	Present	Present	Present	Out-of-synch	Present	2.0–2.2
47	Present	Absent	Present	Out-of-synch	Present	2.0–2.2
48	Absent	Present	Absent	Out-of-synch	Present	2.0– 2.2
49	Absent	Absent	Present	In-synch	Present	2.0–2.2
50	Present	Present	Absent	In-synch	Present	2.0–2.2
51	Absent	Present	Present	Out-of-synch	Absent	2.0
52	Present	Absent	Absent	Out-of-synch	Absent	2.0
53	Present	Present	Present	In-synch	Absent	2.0
54	Absent	Absent	Absent	In-synch	Absent	2.0

The factors used in these simulations are given in Tables S1–4. The model parameters for the detection functions are in Table S5. In order to parameterise the detection functions, the IDCR/SOWER standard line-transect data for seasons from 1978/79–1997/98 were fitted to the equation above, then smoothed to find parameter values yielding distributions of radial sighting distances, sighting angles and perpendicular distances similar to the IDCR/SOWER data (Fig. S2).

6. Output files

The program could output effort and sighting files (Appendix 2 and 3). Data in the sighting output file could either include all surfacings of a group detected by a given platform or just the initial surfacing. In these simulations, only the data from initial sightings were included in the final sighting output files, as this was the general practice in the IDCR/SOWER surveys.

7. Post-processing

This component controlled a variety of options that involved modifying the initial output from the other components above. The software was able to output into the final dataset misidentified duplicate sightings, and to add measurement error to the recorded values of sighting time, radial distance, angle and group size. In addition, the sightings recorded in the final dataset were modified to follow the IDCR/SOWER protocols (see below for details).

In scenarios sc01-54, the sightings from the Bridge were recorded only when the Bridge detected a whale group before one of the other teams. In sc55 from 2012–13, the final dataset also included Bridge team sightings that it detected at the same time as another team.

Table S5a

Model parameter values for the basic detection functions, which might have a group size effect (Syees), weather effect (Wyses) or cue effect. Parameter B4 is in radians (& degrees). Detection Function No. 1: No group size effect (Sno) and a weather effect (Wyses). For weather codes (WC) 0, 1 and 2. Parameter values of B6–B9 are 0.

Model	Team	B1	B2	B3	B4	B5	Detection function number	
Sno-Wyes (WC0)	IO	.0014	1500	4.0	.40 (23°)	-1.75	1.0	
	Top	.0010	1700	3.5	.45 (26°)	-1.75		
	Bridge	.0022	700	6.0	.40 (23°)	-1.75		
Sno-Wyes (WC1)	IO	.0016	1000	4.0	.35 (20°)	-1.75	1.1	
	Top	.0012	1100	3.5	.40 (23°)	-1.75		
	Bridge	.0024	500	6.0	.35 (20°)	-1.75		
Sno-Wyes (WC2)	IO	.0018	700	4.0	.30 (17°)	-1.75	1.2	
	Top	.0016	900	3.5	.35 (20°)	-1.75		
	Bridge	.0026	300	6.0	.30 (17°)	-1.75		

Table S5b

Model parameter values for the basic detection functions, which might have a group size effect (Syees), weather effect (Wyses) or cue effect. Parameter B4 is in radians (& degrees). Detection Function No. 2: A group size effect (Syees) and weather effect (Wyses). For weather codes (WC) 0,1 and 2. Parameter values of B8 and B9 are 0.

Model	Team	B1	B2	B3	B4	B5	B6	B7	Detection Function Number
Syees-Wyes (WC0)	IO	.0018	1700	4.0	.45 (26°)	-1.0	450	0.15	2.0
	Top	.0010	1900	3.5	.50 (29°)	-1.0	450	0.15	
	Bridge	.0022	700 (500)	5.5 (6.0)	.40 (23°)	-1.0	450	0.15	
Syees-Wyes (WC1)	IO	.0022	1100	5.0	.35 (20°)	-1.25	350	0.10	2.1
	Top	.0015	1500	4.5	.40 (23°)	-1.25	350	0.10	
	Bridge	.0024	500	6.0	.35 (20°)	-1.25	350	0.10	
Syees-Wyes (WC2)	IO	.0025	700	6.0	.20 (11°)	-1.50	250	0.05	2.2
	Top	.0020	1100	5.5	.25 (14°)	-1.50	250	0.05	
	Bridge	.0026	300	6.5	.30 (17°)	-1.50	250	0.05	

Table S5c

Model parameter values for the basic detection functions, which might have a group size effect (Syees), weather effect (Wyses) or cue effect. Parameter B4 is in radians (& degrees). Detection Function No. 3: No group size effect (Sno), a weather effect (Wyses) and cue effect. For weather codes (WC0, 0, 1 and 2 Parameter values of B6 and B7 are 0.

Model	Team	B1	B2	B3	B4	B5	B8	B9	Detection Function Number
Sno-Wyes (WC0)	IO	.0014	1500	4.0	.50 (29°)	-1.75	2500	0.01	3.0
	Top	.0010	1700	3.5	.60 (34°)	-1.75	3500	0.01	
	Bridge	.0022	600	5.5	.40 (23°)	-1.75	1700	0.01	
Sno-Wyes (WC1)	IO	.0018	900	4.0	.35 (20°)	-2.00	2000	0.01	3.1
	Top	.0016	1100	4.5	.40 (23°)	-2.00	3000	0.01	
	Bridge	.0024	500	6.0	.35 (20°)	-2.00	1500	0.01	
Sno-Wyes (WC2)	IO	.0022	500	6.0	.20 (11°)	-2.25	1700	0.01	3.2
	Top	.0020	700	5.5	.25 (14°)	-2.25	2500	0.01	
	Bridge	.0026	400	6.5	.20 (11°)	-2.25	1300	0.01	

For some scenarios, measurement errors were added to recorded values of the sighting time, radial distance, angle and group size by modifying the actual values using the following equations with user-specified parameters.

Sighting time measurement errors

In the actual IDCR/SOWER data, errors in sighting time resulted from recording rounded time values. In the simulated data, this was represented by setting the seconds to zero.

Table S5d

Model parameter values for the basic detection functions, which might have a group size effect (Syees), weather effect (Wyes) or cue effect. Parameter B4 is in radians (& degrees). Detection Function No. 4: School size effect (Syees), weather effect (Wyes) and cue effect. For weather codes (WC) 0, 1 and 2.

Model	Team	B1	B2	B3	B4	B5	B6	B7	B8	B9	Detection Function Number
Syees-Wyes (WC0)	IO	.0018	500	4.0	.55 (32°)	-1.5	450	0.15	2000	0.01	4.0
	Top	.0010	750	4.5	.60 (34°)	-1.5	450	0.15	2500	0.01	
	Bridge	.0022	500	6.0	.40 (23°)	-1.5	450	0.15	1500	0.01	
Syees-Wyes (WC1)	IO	.0018	2100	4.0	.55 (32°)	-1.5	0	0	0	0	4.1
	Top	.0010	2300	3.5	.60 (34°)	-1.5	0	0	0	0	
	Bridge	.0022	500	6.0	.40 (23°)	-1.5	0	0	0	0	
Syees-Wyes (WC2)	IO	.0018	500	4.0	.55 (32°)	-1.5	0	0	2500	0.01	4.2
	Top	.0010	750	4.5	.60 (34°)	-1.5	0	0	3000	0.01	
	Bridge	.0022	500	6.0	.40 (23°)	-1.5	0	0	1500	0.01	

Table S5e

Model parameter values for the basic detection functions, which might have a group size effect (Syees), weather effect (Wyes) or cue effect. Parameter B4 is in radians (& degrees). Detection Function No. 5: No school size effect (Sno) and weather effect (Wyes). For weather codes (WC) 0, 1 and 2.

Model	Team	B1	B2	B3	B4	B5	B6	B7	B8	B9	Detection Function Number
Sno-Wyes (WC0)	IO	.0018	1700	4.0	.45 (26°)	-1.0	400	0.15	2000	0.01	5.0
	Top	.0010	1900	3.5	.50 (29°)	-1.0	400	0.15	2200	0.01	
	Bridge	.0022	600	6.0	.40 (23°)	-1.0	400	0.15	1700	0.01	
Sno-Wyes (WC1)	IO	.0022	800	5.0	.35 (20°)	-1.25	375	0.10	1800	0.01	5.1
	Top	.0018	1000	4.5	.40 (23°)	-1.25	375	0.10	2000	0.01	
	Bridge	.0024	500	6.0	.30 (17°)	-1.25	375	0.10	1500	0.01	
Sno-Wyes (WC2)	IO	.0025	500	6.0	.20 (11°)	-1.50	350	0.05	1600	0.01	5.2
	Top	.0022	700	5.5	.25 (11°)	-1.50	350	0.05	1800	0.01	
	Bridge	.0026	400	6.0	.20 (11°)	-1.50	350	0.05	1300	0.01	

Table S5f

Model parameter values for the basic detection functions, which might have a group size effect (Syees), weather effect (Wyes) or cue effect. Parameter B4 is in radians (& degrees). Detection Function: No. 6 only used in sc55: School size effect (Syees), weather effect (Wyes) and no cue effect. For weather codes (WC) 0, 1 and 2. Parameter values of B8 and B9 are 0.

Model	Team	B1	B2	B3	B4	B5	B6	B7	Detection Function Number
Syees-Wyes (WC0)	IO	.0005	-50	2.0	0.8 (46°)	4.2	450	0.15	6.0
	Top	.0009	1900	1.6	1.0 (57°)	4.0	450	0.15	
	Bridge	.0009	1900	1.6	1.0 (57°)	4.0	450	0.15	
Syees-Wyes (WC1)	IO	.0008	-550	3.0	0.4 (23°)	3.2	350	0.10	6.1
	Top	.0012	1400	2.6	0.6 (34°)	3.0	350	0.10	
	Bridge	.0012	1400	2.6	0.6 (34°)	3.0	350	0.10	
Syees-Wyes (WC2)	IO	.0012	-1150	4.0	0.1 (6°)	2.2	250	0.05	6.2
	Top	.0015	900	3.6	0.2 (11°)	2.0	250	0.05	
	Bridge	.0015	900	3.6	0.2 (11°)	2.0	250	0.05	

Radial distance and sighting angle measurement errors

Errors in radial distance and sighting angle were generated together by initially including sightings detected at angles up to 120°, since some of these could be reduced to $\leq 90^\circ$ after errors were incorporated. After measurement error was added into the angles, sightings that resulted in 'recorded' angles greater than 90° were discarded from the final sightings files.

Errors in radial distance and angle were generated according to the following models, as specified in Skaug *et al.* (2002):

$$r_{err} = 0.898r \cdot \exp(1.371 \cdot Normal(0,1) \cdot r^{-0.2375})$$

$$\theta_{err} = 1.057\theta + [4.826 \cdot Normal(0,1) \cdot \exp(0.0117 \cdot \theta_{55})]$$

where r = radial distance without error,

r_{err} = radial distance with error

θ = angle without error

θ_{err} = angle with error

θ_{55} = minimum of the $\text{abs}(\theta)$ or 55°

$Normal(\mu, \sigma^2)$ = a normal distribution with mean = μ and variance = σ^2

The effect of these models was that the magnitude of error on the angle (Fig. S3), radial distance (Fig. S4) and perpendicular distance (Fig. S5) all increased as distance from the track line increased, particularly the radial distance.

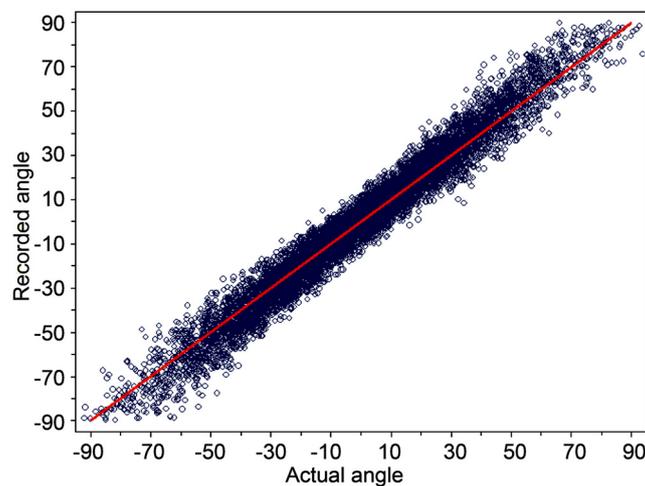


Figure S3. Effect of measurement error on the recorded angles (as measured in degrees). An angle of zero represents a sighting on the track line, while negative values represent sightings made on the left side of the track line (port-side). Red line represents the 1:1 relationship.

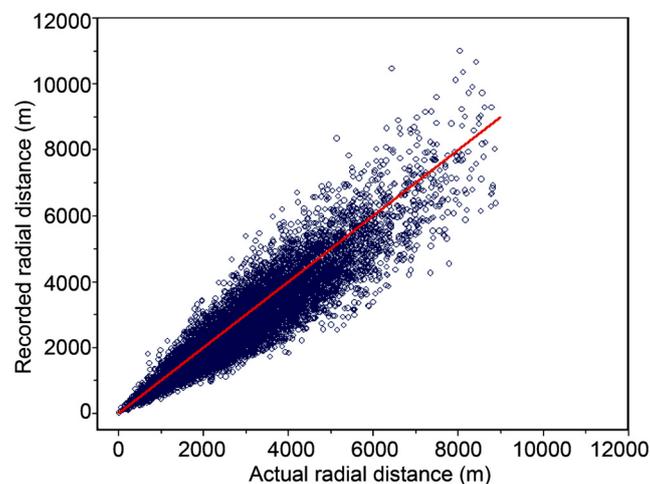


Figure S4. Effect of measurement error on the recorded radial distances (as measured in meters). A radial distance of zero represents a sighting next to the ship. Red line represents the 1:1 relationship.

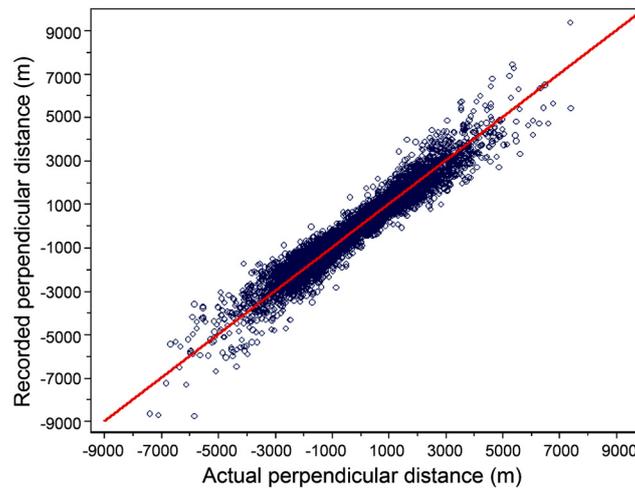


Figure S5. Effect of measurement error on the recorded perpendicular distances (as measured in meters). A perpendicular distance of zero represents a sighting seen right in front of the ship on the track line, while negative values represent sightings made on the left side of the track line (port-side). Red line represents the 1:1 relationship.

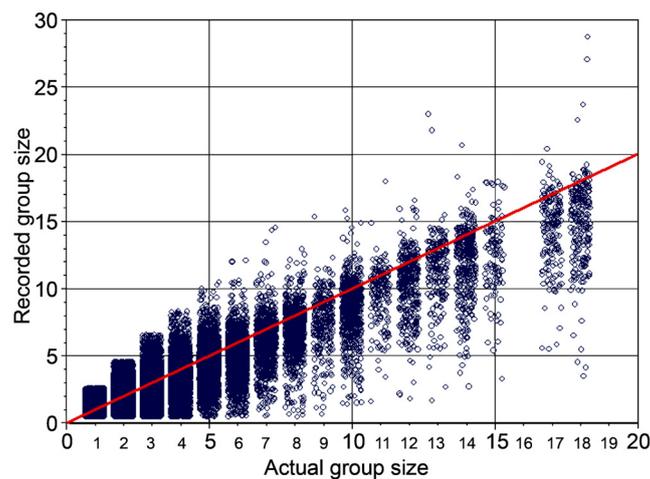


Figure S6. Actual and recorded group sizes in simulation scenarios with group-size measurement error, jittered to view better. Red line represents the 1:1 relationship.

Group size measurement errors

As a simplification, no measurement error was added to the recorded group size in the initial simulation scenarios sc01-16 (2004), so all group sizes recorded in either Closing or IO mode were correct. In some scenarios after sc16, a more realistic situation was simulated where group-size measurement error was present. When group-size measurement error was present, it was assumed that only group size estimates for sightings closed on were correct, so no measurement error was added to these sightings. Sightings closed on were those detected in Closing mode and located at 0.3–3.0 n.mi perpendicular distance from the track line (with error, if any). Consequently, when measurement error was present, measurement error was added to all sightings detected in IO mode and sightings detected in Closing mode that were less than 0.3 n.mi or more than 3.0 n.mi from the track line.

The added measurement error resulted in sizes of groups that could be greater or less than the actual group size, but always at least one. Most of the group sizes with error were less than the corresponding actual group size (Fig. S6). The probability of detecting a group depended on the actual group size, not the group size recorded with error.

Table S6
Percentage of total groups with a specified true group size recorded as various group sizes

True size	Estimated group size					Total
	1	2	3–4	5–9	10+	
1	95	5	0	0	0	100
2	55	43	2	0	0	100
3–4	23	32	39	6	0	100
5–9	5	3	45	47	0	100
10+	0	0	22	33	45	100

The function to generate group sizes with error had four components. The first component was a base estimate of group size, which depended on the perpendicular distance of the sighting, and in the 2012–13 scenarios, ranged from the actual group size at perpendicular distances of 0–20% of the actual group size at very large perpendicular distances. The second component caused the variance of the error to increase with larger group sizes. The third component caused the variance of the error to increase with larger perpendicular distances. In addition, for scenario sc55 (2012–13), a fourth component was added: a normally distributed random element (details below). Parameter estimates for the sc55 scenario approximated the patterns documented in the SOWER group size experiment (Hedley, 2012) (Table S6).

For sc55 (2012–13), the four error-addition components were as follows (with the result also constrained by limits on the minimum and maximum group size when error was added):

1. An ‘S-curve’ factor that determined the level of under-estimation depended on the location of the group. This was set so that the base under-estimation declined for all of the true group sizes at perpendicular distances of zero toward 20% of the true group size at large perpendicular distances. The base estimate was:

$$base = 0.2 + \frac{\text{invlogit}[-f_1 \cdot (perp - errorDist - f_2)]}{0.8 \cdot \text{invlogit}(f_1 \cdot f_2)}$$

where:

$$\text{invlogit}(x) = \frac{\exp(x)}{\exp(x) + 1}$$

and $base$ = base under-estimate,

$perp$ = perpendicular distance

$errorDist$ = minimum error distance

f_1, f_2 = parameters for tuning shape of function.

The 0.8 in the denominator caused the value of the denominator to range between 0 and 0.8 rather than 0 and 1. The values used were $f_1 = 0.0040$ and $f_2 = \text{minimum Closing distance} = 0.3 \text{ n.mi or } 555.6 \text{ m}$.

2. A factor that caused the variance of error magnitudes around the base estimate to increase with greater group size (see first factor in right-hand side of the *randfactor* equation below).
3. A normally distributed random component (second factor on right-hand side of the *randfactor* equation):

$$randfactor = \left(\frac{3 \cdot origGS}{15 + origGS} + 0.2 \right) \cdot Normal(0, 1)$$

where $origGS$ was the original group size without error and $Normal(\mu, \sigma^2)$ was a normal distribution with mean = μ and variance = σ^2 .

4. A component that caused the variance of error magnitudes around the base estimate to increase with greater perpendicular distance.

Thus, the group size with error, GS_{err} , was defined as:

$$GS_{err} = \text{anint} \left[(base \cdot origGS) + \left(randfactor \cdot \left(\frac{perp + 500 - errorDist}{2000} \right) \right) \right]$$

where anint is a function that rounds to the nearest integer.

Finally, GS_{err} was subject to the constraints that (a) $1 \leq GS_{err} \leq (2 \cdot origGS)$, and (b) $GS_{err} < 3$ is not permitted if $origGS \geq 10$.

If a value for GS_{err} violated these constraints, a new value was generated by resampling. If, after 100 tries, no satisfactory value was generated, the process was then assumed to be 'stuck,' and if the absolute limits above had been exceeded, the value of GS_{err} was changed to 1 or to $(2 \cdot origGS)$, whichever was nearest to the last value generated. However, cases in which GS_{err} values were actually forced to 'pile up' at the limits in this way were rare. If the last value generated had $GS_{err} < 3$ while $origGS \geq 10$, GS_{err} was set to 3.

In general, this error model was intended to generate errors tending more to the lower side than in earlier scenarios (see below for earlier versions).

For scenarios sc33-54, group size error was generated as above, except that the equation for the base [under]estimate was:

$$base = 0.75 + \frac{\text{invlogit}[-f_1 \cdot (perp - error \cdot dist - f_2)]}{4.0 \cdot \text{invlogit}(f_1 \cdot f_2)}$$

and the equation for the random component was:

$$randfactor = \left(\frac{3 \cdot origGS}{15 + origGS} + 0.7 \right) \cdot Normal(0,1)$$

Furthermore, there was no constraint preventing groups of true size ≥ 10 from being estimated as < 3 .

For scenarios sc17-32, a slightly different function was used to generate group size errors: the base estimate factor ('base' in the equations above) was set to a constant 0.8. In addition, 500 was not added in the numerator of the fraction involving $perp$. In addition, group size errors were not permitted for groups closer than the minimum perpendicular distance for Closing (' $minclosing$ '), so that this distance (0.3 n.mi) was functionally equivalent to $errorDist$.

Thus:

$$GS_{err} = \text{anint} \left[(0.8 \cdot origGS) + \left(randfactor \cdot \left(\frac{perp - minclosing}{2000} \right) \right) \right]$$

where $minclosing$ was the minimum Closing distance, which is the minimum error distance.

This treatment of group errors led to a sharp and seemingly unrealistic change in the modal group size estimate just beyond the minimum distance for Closing in these scenarios – also the minimum distance for error; hence, the error formula was adjusted before use in later scenarios, as described above.

Bridge data

For scenarios sc01-54, during IO mode, it was possible for a group to be seen and recorded by all three teams: the Topman, IO and Bridge teams. However, Bridge team sightings were recorded in the final output sighting file only when the Bridge team detected a group before either the Topman or IO team detected the same whale group. In other words, if the Bridge team initially detected a surfacing of a whale group at the same time or after

either the Topman or IO teams did, then the Bridge team did not record this whale group. For scenario sc55, this restriction was removed, so when the Bridge team detected a group at the same time as either or both the other teams, this Bridge team sighting was recorded in the final dataset.

In addition, for scenarios sc01-54, the Bridge recorded no sightings during any period for which any group seen by any platform had not yet passed the ship. For scenario sc55, this constraint was eliminated.

Sighting modes

In IO mode, no survey time was lost to ‘closing’ on groups. Thus, there were no gaps in the survey effort along the track line. During scenarios in which IO and Closing mode alternated, each mode lasted approximately four hours, with the starting time for each mode indicated in the Effort file. During Closing mode, when a group was detected within 3 n.mi of the track line but more than 0.3 n.mi away from it, the ship went ‘off effort’ to head to the group and confirm its size (time and position indicated in the Effort file), before returning to the track line (time and position where it rejoined the track line again indicated in the Effort file). In these simulations, search effort (the ‘on effort’ condition) did not resume until the ship was back on the original track line. This is in contrast to the IDCR/SOWER protocol where search effort started up again at the point when the group visited was left and the ship started heading back to the original track line. Because of time spent ‘off effort’ during Closing mode in these simulations, there were gaps in the track line that were not surveyed. No groups that could only have been seen in these ‘off effort’ gaps were recorded in the final dataset.

In summary, consistent with the IDCR/SOWER protocols, the term ‘confirmed group size’ refers to sizes of groups that were either closed on (so the group size was correctly recorded) or were within 0.3 n.mi of the track line (in either Closing or IO mode). Note that in some scenarios error could have been added to the recorded size of groups within 0.3 n.mi of the track line, since it is assumed that only the size of groups closed on were correct.

If, when the ship returned to on-effort from a closing, the time was past that scheduled for the start of IO mode, then IO started at whatever time on-effort started and lasted until the next normal time to switch. For example, normally each mode would last four hours. Normally, the modes would change at 00:00:00, 04:00:00, 08:00:00, 12:00:00, etc. So, if Closing mode started at 04:00:00, IO mode would normally start at 08:00:00. If there was a sighting to be closed on at 07:50:00 and the ship only returned to the track line at 08:06:21, then on-effort would resume and IO mode would start at 08:06:21. Closing mode would start again at 12:00:00, even though IO mode was not exactly four hours long.

Duplicate sighting misidentification

If two or more groups were close enough in time, space and group size, one could be mistaken for another. In practice, this happened relatively rarely as the densities assumed for the simulations were so low (roughly 0.1–0.3% of total sightings in final output).

Criteria applied for groups to potentially be confused were as follows:

1. True group sizes of the respective groups differed by no more than a factor of two;
2. Sightings of the two groups were within 20 minutes of each other; and
3. Distance between the two sightings was less than a ‘maximum confusion distance,’ currently set at 25% of the mean of the radial distances to the two sightings.

If all these criteria were met, confusion of a given group with any other group within its ‘confusion range’ could occur. This actually occurred randomly at a frequency that depended on how many total groups were within this confusion range. In the most common case, where there was only one other group that was a confusion candidate, confusion occurred about 50% of the time.

The procedure to simulate misidentifying duplicates started with all surfacings of all sightings, not just the first surfacings of each sighting that were recorded in the final output. Ten percent of sightings were given incorrect group ID numbers. The new group numbers were drawn from those not already recorded for that day in the first-pass output. This allowed for two types of misidentification of duplicates or triplicates. First, for actual different

groups in the same general vicinity that were detected by multiple platforms, the misidentification could result in one of these groups being recorded as a duplicate or triplicate of the other group but from a different surfacing of the group. Second, for an actual group of animals that was truly a duplicate or triplicate detected by multiple platforms, the misidentification of the group that surfaced multiple times could result in one surfacing being recorded as a group detected by one platform, but another surfacing of the same group being recorded as a different new group by the other platform(s). This is why there were 12–14% altered group ID's in final output, when only 10% were initially set as altered. Note there were two kinds of changes that did not affect the density: (a) when a group was seen only once, and its ID was changed through the random process; (b) when the confusion process happened to be applied in a reciprocal fashion; i.e., where Group A was given Group B's ID while Group B was given Group A's ID and no other sightings of these groups remained in the final output.

The chance of assigning a new ID by mistake did not depend on the time or distance since the last previous sighting of the group in question.

METHODS TO CALCULATE 'TRUE' VALUES

C-shell and Matlab scripts were used to determine the realised density of whale groups and mean group size over the viewable area that the survey ship actually covered in a given simulation. These scripts tallied potentially sightable whale groups by first obtaining the maximum viewing distance reached by the ship each day (coordinate of [maximum distance reached by ship on track + maximum sighting distance]) and the ship's maximum sighting radius strip along the track line, then obtaining positions and sizes of whale groups actually 'seeded' within this simulation area.

The values of $g(0)$, the probability of detecting a group on the track line, and effective-strip-half-width (ESHW) were determined only for runs entirely in IO mode. In order to obtain $g(0)$ and ESHW values for scenarios that included Closing mode, post-processing was repeated to generate realisations with the IO mode used throughout. In $g(0)$ and ESHW determination, neither the 9,000 m at the start of the ship's daily track nor the 9,000 m beyond the point where the ship stopped each day were considered. This was because whale groups in these areas did not pass through the ship's full 'detection window'.

The $g(0)$ value was defined as the ratio of the total number of whale groups detected by a given platform within a given small distance from the track line to the total number of groups positioned within that narrow strip at the start of each model day's run. Locations for these groups were obtained from the simulation's whale configuration files.

The maximum distance from the track line used was 20 m when $g(0)$ was derived from the pooled 100 replicates of the 'regular' runs for each scenario. For the special-purpose runs where all whale groups were pre-positioned on the track line, $g(0)$ was derived from the numbers detected on the track line. Whales positioned in space before the ship's daily starting position or after the ship's daily ending position were not counted in either numerator or denominator of the $g(0)$ ratio, even though some of those beyond the ending position may have been detected.

The ESHW value was determined by using the built-in Matlab minimisation procedure 'fminsearch' to look for the perpendicular distance from the track line such that the total number of whales detected beyond it equaled the total number of whales not detected within it. Location data from the whale configuration file(s) were used to establish how many whales were actually there.

REFERENCES

- Branch, T.A., & Butterworth, D.S. (2001). Southern Hemisphere minke whales: Standardised abundance estimates from the 1978/79 to 1997/98 IDCR-SOWER surveys. *J. Cetacean Res. Manage.* 3(2): 143–74. [Available at: <https://doi.org/10.47536/jcrm.v3i2.886>]
- Hedley, S., & Ensor, P.H. (2006). A preliminary look at minke whale dive time data from the 2004/05 IWC-SOWER survey. SC/58/IA21 presented to the IWC Scientific Committee, St. Kitts and Nevis, 2006. [Available from the IWC Publications Team]
- Hedley, S.L. (2012). Respiration rates of Antarctic minke whales from the Visual Dive Time Experiment on the 2004/05 IWC SOWER Survey. SC/64/IA12 presented to the IWC Scientific Committee, Panama, 2012. [Available from the IWC Publications Team]
- Murase, H., Palka, D., Punt, A.E., Pastene, L.A., Kitakado, T., Matsuoka, K., Hakamada, T., Okamura, H., Bando, T., Tamura, T., Konishi, K., Yasunaga, G., Isoda, T., & Kato, H. (2020). Review of the assessment of two stocks of Antarctic minke whales (eastern Indian Ocean and western South Pacific). *J. Cetacean Res. Manage.* 21: 95–122. [Available at: <https://doi.org/10.47536/jcrm.v21i1.181>]

- Popenoe, H. (2000). Modifications to the TRANSIM program that allow reactive movement. Contract report. Available from D. Palka, NMFS, 166 Water St., Woods Hole, MA 02543.
- Skaug, H.J., Oien, N., Boethun, G., & Schweder, T. (2002). Abundance of northeastern Atlantic minke whales for the survey period 1996–2001. SC/54/RMP5 presented to the IWC Scientific Committee, Shimonoseki, 2002. [Available from the IWC Publications Team]
- Ward, A.J. (1988). Accounting for the effect of pod size in the estimation of blow rates: Analysis of data from the 1984/85, 1985/86 and 1986/87 IDCR minke whale assessment cruises. *Rep. Int. Whal. Comm.* 38: 265–8.

©Authors. This is an open access article distributed under the terms of a *Creative Commons License CC-BY-NC 4.0*.

Appendix 1

DESCRIPTION OF SIMULATED REACTIVE BEHAVIOUR

Overview

Reactive behaviour in the simulation program occurred when a whale entered into a reactive region within a user-defined distance from the ship (*ReactionRadius*), then reacted with changes in dive time, speed and/or direction, resulting in an altered surfacing position. The amount of reaction was a function of the distance between the whale and ship and the user-defined reactive distance. As taken from Popenoe (2000), the formulas below summarise the reaction factor, dive time, speed and direction.

Reaction Factor

There were separate reaction factors for dive time, speed, and direction respectively, but all used the following general function with different *Alpha*'s and *Beta*'s for each of Dive, Speed, or Direction:

$$ReactionFactor = \left(\frac{ReactionRadius - Z}{ReactionRadius} \right)^{Alpha} \left(\frac{ReactionRadius - X}{ReactionRadius} \right)^{Beta}$$

ReactionRadius = User-defined maximum radial distance within which reactive behaviour was initiated;

Z = Distance between ship whale at time NOW;

X = Distance between ship and whale at NEXT surfacing, with whale having no reaction;

AlphaDive = User-defined scaling factor;

AlphaSpeed = User-defined scaling factor;

AlphaDirection = User-defined scaling factor;

BetaDive = User-defined scaling factor;

BetaSpeed = User-defined scaling factor;

BetaDirection = User-defined scaling factor;

If the *ReactionFactor* came out negative, then the *ReactionFactor* was set to zero.

Dive Time

The 'base' dive time (disregarding reaction) was determined by the program's 'surface' function. The equation for the realised dive time that included reaction was:

$$\begin{aligned} RealisedDiveTime &= DiveTime + [(MaxDive - DiveTime) \cdot DiveReactionFactor] \\ &+ [Normal(0, DiveVariance) \cdot DiveReactionFactor] \end{aligned}$$

where:

DiveTime = dive time of the whale without reactive behaviour;

MaxDive = maximum dive time of a whale (user-defined);

Normal = the normal distribution function;

DiveVariance = variance for the normal distribution function, where $DiveVariance = (StdDevDive)^2$ and *StdDevDive* is user-defined.

If *RealisedDiveTime* exceeded *MaxDive*, it was set equal to *MaxDive*.

Speed

The equation for the realised speed was as follows:

$$\begin{aligned} \text{RealisedSpeed} &= \text{Speed} + [(\text{MaxSpeed} - \text{Speed}) \cdot \text{SpeedReactionFactor}] \\ &+ [\text{Normal}(0, \text{SpeedVariance}) \cdot \text{SpeedReactionFactor}] \end{aligned}$$

where:

Speed = speed of a whale without reactive behaviour;

MaxSpeed = maximum speed of a whale (user-defined);

SpeedVariance = variance for the normal distribution function, where $\text{SpeedVariance} = (\text{StdDevSpeed})^2$ and StdDevSpeed is user-defined.

If *RealisedSpeed* was greater than *MaxSpeed*, then *RealisedSpeed* was set equal to *MaxSpeed*. If *RealisedSpeed* came out less than the original *Speed*, then *RealisedSpeed* was set equal to *Speed*.

Direction

Angles were considered to be 0 at a positive X-axis direction (i.e., at 3:00 position on a clock), with relative positive angles in a counter-clockwise direction. The direction reaction behaved in a manner such that, if a whale was on the right side of the ship, the whale's direction was changed towards a zero angle (0 radians), that is, the 3:00 direction. Likewise, if the whale was on the left side of the ship, the whale's direction was changed toward a 180° (π radians), that is, the 9:00 direction.

In the program, the direction of a whale was a normalised vector of an *Xvelocity* and *Yvelocity*. The angle of direction was determined by:

$$\text{WhaleDirection} = \arctan\left(\frac{\text{Yvelocity}}{\text{Xvelocity}}\right)$$

The equation for the change in direction was as follows:

$$\begin{aligned} \text{RealisedGamma} &= \text{NowGamma} - (\text{Beta} \cdot \text{DirectionReactionFactor}) \\ &+ [\text{Normal}(0, \text{DirectionVariance}) \cdot \text{DirectionReactionFactor}] \end{aligned}$$

where:

NowGamma = direction of the whale without reactive behaviour;

Beta = angle between: (1) a line perpendicular to and pointing away from the ship's track line; and (2) the whale's heading vector;

DirectionVariance = variance for the normal distribution function, where $\text{DirectionVariance} = (\text{StdDevDirection})^2$ and StdDevDirection is user-defined.

When the whale was on the left side of the ship, $\text{Beta} = \pi - \text{abs}(\text{NowGamma})$. In addition, the minus sign in the above equation for *RealisedGamma* became a plus sign when the whale was on the left side. When the whale was on the right side of the ship, $\text{Beta} = \text{abs}(\text{NowGamma})$.

Realised Position

The calculation for the realised position was as follows:

$$\text{RealisedPosition} = \text{PreviousPosition} + (\text{RealisedSpeed} \cdot \text{RealisedGamma} \cdot \text{RealisedDiveTime})$$

where *RealisedSpeed*, *RealisedGamma* and *RealisedDiveTime* were the results from the above equations.

Appendix 2

EFFORT DATA FORMAT

Each line in the effort file provided the time at which the circumstances of the sighting process changed: e.g., the time when a new watch team started, when the weather changed, when IO mode started after being in Closing mode, or when surveying ended.

The fields in the effort file included:

1. Year
2. Month
3. Day
4. Hour:Minute:Second
5. Platform number
6. Watch code
7. Weather code
8. Survey mode code
9. X position
10. Y position

The codes used in these fields were:

- Platform codes: 0 (IO platform), 1 (Topman), 2 (Bridge), -1 (no change from before)
- Team codes: 0 (first team on a platform), 1 (second team on the same platform), -1 (no change from before)
- Weather codes: 0 (Beaufort 0), 1 (Beaufort 1–2), 2 (Beaufort 3 and above), -1 (no change from before)
- Mode codes: I (IO mode), C (Closing mode), N (going on-effort during Closing mode now), F (going off-effort during Closing mode now), E (end of effort), W (initial weather conditions for the day or weather conditions change now)

Every day of surveying started at time 0:0:0 (stated in file) and ended at 16:00:00, except in the 2008 scenarios when days might end later.

The effort file contained one row for each circumstance that changed. In the following example, the initial weather was indicated in the first line; the starting mode (IO) was indicated in the second line; and the next three lines indicated the starting team of observers for each platform. Thus, in this example, on Month = 1, Day = 1, Year = 1995, the initial weather was code 0; surveying started at 0:0:0 in IO mode; and each platform (0, 1 and 2) started with Team 0. Two hours later, at 2:0:0, the observer teams on each platform switched, so Team 1 was now on duty and Team 0 was off duty. Since IO mode had not changed, there was no row for such a change at time 2:0:0.

As a side-effect of the processing that generated these files, lines specifying which teams were on watch might occasionally appear when the teams were not actually changing. These lines correctly specify the teams that continued on watch at these times.

Example of the effort data (with column headings added) follows:

Table S7

Year	Mon	Day	hh:mm:ss	Platform	Watch	Weather	Mode	Xposn	Yposn
1995	1	1	0:00:00	-1	-1	0	W	0	18532
1995	1	1	0:00:00	-1	-1	-1	I	0	18532
1995	1	1	0:00:00	0	0	0	I	0	18532
1995	1	1	0:00:00	1	0	0	I	0	18532
1995	1	1	0:00:00	2	0	0	I	0	18532
1995	1	1	2:00:00	0	1	0	I	0	61157
1995	1	1	2:00:00	1	1	0	I	0	61157
1995	1	1	2:00:00	2	1	0	I	0	61157

Appendix 3

SIGHTING DATA FORMAT

The fields in the sighting data files were:

1. Year
2. Month
3. Day
4. Hour:minute:second
5. Platform code (0 = IO, 1 = Topman, 2 = Bridge)
6. Whale group ID number
7. Distance ahead along the track line
8. Perpendicular distance between the track line and location of the center of the whale group
9. Radial distance between the locations of the ship and whale group
10. Relative angle between the track line and line of sight to the whale group
11. Relative swim direction of whale group
12. Group size (if 0, assume 1 whale)
13. Observer team code (0 or 1)
14. Weather code (0, 1, or 2)
15. Cue code (0 = body, 1 = blow)
16. X position of ship at time of sighting (m)
17. Y position of ship at time of sighting (m)
18. Mode code (0 = IO mode, 1 = Closing mode)
19. Group size confirmation status code (0 = group size not confirmed, 1 = group size confirmed)

The first 14 variables (year to weather code) were available for all years when data were simulated. Variables 15 to 18 (cue to mode) were available for 2005 and thereafter. Variable 19 (group size confirmation) became available in 2008.

Relative sighting angles and swim directions ranged from 0–360°. For sighting angles, 0° was assigned to a group that was directly ahead on the track line, 90° to a group directly abeam on the right-hand (starboard) side of the ship, 180° to a group on the track line but behind the ship, and 270° to a group directly abeam on the left-hand (port) side of the ship.

Swim directions followed the same circular assignment but indicated the direction in which a group was swimming towards. For example, a swim direction of 0° indicated that a group was swimming in the same direction as the ship. A swim direction of 270° indicated that a group was swimming in a right-to-left direction, no matter if the group was on the right or left side of the ship.

Perpendicular distances were either positive or negative. Negative perpendicular distances indicate a sighting was on the left side of the ship, i.e., where the sighting angles were greater than 180°.

Group sizes recorded in this file were the same for all teams/platforms that detected a given group, except in scenarios where recorded group sizes could have errors. Sizes for a given group could be different for different platforms in cases where initial detections of that group did not trigger a closing and a group size recording error occurred, but a subsequent detection of that group by a different platform did trigger that group to be closed upon. After the group had been closed upon, the group size recorded was the correct group size.

Example of sighting data:

Table S8

Year	Mon	Day	hh:mm:ss	Platform	ID	Ahead	Perp	Radial	Angle	Swim direction	Group size	Watch	Weather	Cue	X	Y	Mode	Confirmed
1995	1	1	1:14:00	1	94	2503.93	1629.40	2987.41	33.05	212.51	2	0	0	0	0	45043	0	0
1995	1	1	1:16:00	1	184	1073.34	384.11	1140:00	19.59	348.59	2	0	0	1	0	45866	0	1
1995	1	1	1:18:00	0	184	721.11	428.91	838.91	30.73	348.59	2	0	0	1	0	46552	0	1