Comparison of results from the OK, SPLINTR, Integrated and standard analytical methods when applied to simulated data, 2004 – 2008

Debra Palka
Northeast Fisheries Science Center, Woods Hole, MA 02543 USA

INTRODUCTION

The IWC/SOWER abundance line-transect survey data has been analyzed by several newly developed methods and by the "standard analysis method". The International Whaling Commission (IWC) Scientific Committee (SC) recommended simulated line transect data be used to evaluate the robustness of these analysis methods. To do so, simulated line transect data were generated that incorporated biases due to heterogeneity in factors related to the distribution, density and behavior of minke whales and to the manner in which the sighting surveys were conducted.

The 2004 set of simulated scenarios (sc01 to sc16) incorporated effects of the following factors (Table 1A):

1) Whale density follows a horizontal gradient
2) School size follows a horizontal gradient
3) Whale groups are clustered
4) Survey data are collected under the procedure of IO only or alternating between IO and Closing
5) Probability of detecting a group depends on group size
6) Probability of detecting a group depends on weather conditions
7) Probability of detecting a group depends on the initial cue.

The 2005 set of simulated scenarios (sc17 to sc32) incorporated effects of the following factors (Table 1B):

1) Recorded data includes school size errors
2) Whale density is correlated with weather sighting conditions
3) Whale density has a vertical and horizontal density gradient
4) Survey data are collected under the procedure of IO only or alternating between IO and Closing
5) Individual whales within a group surface in either a synchronized or non-synchronized fashion.

The 2006 and 2007 set of simulated scenarios (sc33 to sc38) incorporated effects of the following factors (Table 1C):

1) Recorded data includes schools size errors, more than that used in 2005
2) Probability of detecting a group depends on weather conditions
3) Whale density follows a gradient that could be monotonic or non-monotonic
4) Whale density has a horizontal and vertical gradient
5) School size follows a gradient
6) Individual whales within a group surface in a non-synchronized fashion
7) Recorded data includes errors in the radial distance and angles

The 2008 simulated scenarios (sc39 to sc54) incorporated effects of the following factors (Table 1D):

1) Recorded data includes measurement errors in the time, radial distance, angle and school size
2) Whale density follows a non-linear horiontal gradient
3) Some groups are incorrectly assigned as a duplicate sighting
4) Individual whales within a group surface in a non-synchronized fashion
5) Location correlation. That is, simultaneously, there is a school size gradient (larger schools near ice), vertical density gradient (higher density near ice and non-linear) and weather gradient (better weather near ice), so the detection function is dependent on school size and weather.

To investigate the robustness of an analytical method to the above factors, values of the actual simulated animal density were compared to values of the estimated animal density resulting from the analytical methods. The analysis methods investigated were the standard analysis method (standard) by Branch, hazard probability method (OK) by Okamura and Kitakado, SPLINTR method by Bravington and Hedley, and the integrated model method (Int) by Cooke. During the last few years there have been several versions of the OK model.

Results from all the versions are presented in Tables 3-5, though only the most recent version (OK-2009) will be discussed in the results section.

METHODS

To investigate the effects of the factors and to de-alias the 2-way and higher interactions from the main effects of these factors, a Mirror-Image Fold-over partial factorial design was developed. Thus, scenarios created in 2004, 2005, and 2008 were designed to be investigated as a set because each scenario included a different combination of the factors. A few more simulations were created in 2006 and 2007, where the scenarios included different combinations of factors and a couple of additional factors that had been suggested by the Committee in previous meetings.

To quantify and determine the significance of the effect of these factors, the percent bias, (observed – actual))/actual, was calculated for the mean of the 100 replications within each scenario.

2004

For the 2004 scenarios, to investigate which effects corresponded with less accurate density estimates, an ANOVA was conducted, where the model, as specified in SPLUS language, was:

Mean percent bias ~ dens.grad + size.grad + clump + mode change + df.size + df.weather + df.cue

where each factor on the right-hand side of the equation was a dummy variable:

–1: indicating a factor is not present (whale groups randomly distributed, group sizes randomly distributed, whale groups randomly distributed, only IO mode, probability of detecting a group is not dependent on the group size, weather conditions, or initial cue), or

+1: indicating a factor is present (whale density is highest at the ice edge and lowest farthest from the ice edge, group sizes are the highest at the ice edge and smallest farthest from the ice edge, whale groups are spatially clustered together, Closing and IO modes are alternated, and probability of detecting a group is dependent on the group size, weather conditions or the initial cue).

2005

For the 2005 scenarios, the ANOVA model used to investigate the robust of the estimated abundances, as specified in SPLUS language, was:

Mean percent bias ~ size.errors + weather.density + vert.horz.density + mode.change + non.synch

where each factor on the right hand side of the equation was a dummy variable:

–1: indicating a factor is not present (no size errors, no correlation between weather and whale density, no vertical and horizontal whale density, only IO mode, and whales surface in total synchrony), or

+1: indicating a factor is present (errors are in the recorded group size, whale density is correlated to weather sighting condition, whale density follows a vertical and horizontal gradient within the study area, Closing and IO mode are alternated, and individual whales surfacing patterns are not synchronized).

2008

For the 2008 scenarios, the ANOVA model used to investigate the robustness of the estimated abundances, as specified in SPLUS language, was:

Mean percent bias ~ size.errors + horz.density + mis.id + non.synch + location.correlation

where each factor on the right-hand side of the equation was a dummy variable:

–1: indicating a factor is not present (no measurement errors, no horizontal density gradient, no mis-identification of duplicate sightings, whales surface in total synchrony, and no vertical whale density, no school size gradient, no weather gradient), or

+1: indicating a factor is present (measurement errors are in the recorded values, whale density follows a horizontal gradient within the study area, there is some mis-identification of duplicate sightings, individual

whales surfacing patterns are not synchronized, and the following simultaneously: whale density follows a non-linear vertical gradient, school sizes follow a horizontal gradient, weather conditions change over the day and influence the probability of detecting a group of whales, and school size influences the probability of detecting a group of whales).

RESULTS

The actual simulated animal density for scenarios sc01 to sc16 ranged from about 0.058 to 0.071individual whales per km^2; for scenarios sc17 to sc32, densities were about 0.072 individuals per km^2; for scenarios sc33 to sc38, densities ranged from 0.029 to 0.107 individuals per km^2; and for scenarios sc39 to sc54 densities ranged from about 0.058 to 0.102 individuals per km^2 (Table 2).

The results focus on the OK-2009 and SPL-2009 methods as compared to the standard method.

Examining the distribution of the percent bias of each of the iterations within a scenario the following observations were made (Figures 1-3):

1. In general, estimates of SPL were slightly less bias than those from OK in the complex scenarios (sc39-54). In contrast, SPL was slightly more negatively biased than the OK estimates in the less complicated scenarios (1-16). In this case not biased is defined as the actual value being within the center half of the distribution of density estimates (blue boxes in box plots).

2. In the complex scenarios, without mis-identification duplicates, for the OK method the estimates of density of schools were generally negatively biased and the estimates of group size were positively biased, resulting in a mixture of positively and negatively or not biased estimates of whale density. The SPL method also does some such compensations but to a smaller degree. That is, the estimates of group size were generally slightly positively biased, estimates of school density was either negatively biased or not biased, resulting in estimates of whale density that were either not biased or slightly negatively biased.

Examining the differences between the results from the OK and SPLINTR method when comparing both results of analyzing the same data within an iteration (Figures 4-6):

3. There was a lot of variation between the results obtained by the two methods even when analyzing the exact same data from the same iteration for the scenarios with the complicated location correlation [simultaneously whale density follows a non-linear vertical gradient (i.e., across days or along the ice edge), school sizes follows a horizontal gradient (i.e., within a day or perpendicular to the ice edge), weather changes within a day (perpendicular to the ice edge), and the detection probability was influenced by group size and weather], scenarios 43-50.

4. Most of the results from OK are greater than that from SPL when estimating the density of schools and whales. However, the estimates of group size were more similar for the two methods. The scenarios where the estimates from OK were greater than SPL were in general scenarios with un-modeled heterogeneity. More specifically the following factors:

    a. When cue influenced the detection function

    b. When there was the complex location correlation [simultaneously whale density follows a non-linear vertical gradient (i.e., across days or along the ice edge), school sizes follows a horizontal gradient (i.e., within a day or perpendicular to the ice edge), weather changes within a day (perpendicular to the ice edge), and the detection probability was influenced by group size and weather], surfacing whales within a group were in-sync with each other, and there was mis-identification of duplicates, and

    c. When there was non-synchronized diving. (This could be interpreted as un-modeled heterogeneity. That is, what is happening in the simulation is as a group is surfacing in front of the ship, some or all of the group surfaces, the probability of detecting that surfacing is dependent on the number of animals that surfaced, but the recorded number of animals in the group is the total number of animals in the group (with or without measurement error). The logic of recording the total number of animals and not the number from that particular surfacing was the data collection process is the following: teams watch the group as it surfaces, often multiple times, and then the team assesses (and thus records) the total number of animals in the group.)

When examining the comparison of the means of the 100 scenarios within each scenario that did not have mis-identified duplicates, the following observations were made:

5. The estimates of mean group sizes by all three methods were similar (Table 2). The average (and median) ranged from 2-6 percent bias, where the OK and SPL methods had similar levels of variability of the percent bias (RMS≈6.5) and the standard method was more variable (RMS=9.7).

6. The estimate of school density for the SPL method was nearly always negatively biased (average=-9.5%, median=-9.6%, RMS=10.3). In contrast, the school density for the OK method was positively biased to a small absolute degree and slightly less variable (average = +0.8%, median=+2.2%, RMS=7.2). The standard method was the most negatively biased and more variable (average=-19.6, median=-19.7, RMS=21.4).

7. The estimates of whale density for all three methods follow the same patterns as the school density estimate, since the whale density is essentially the product of the school density and the mean group size.

8. From the scenarios that incorporate many of the complicating factors found in the IDCR/SOWER minke whale abundance data (measurement error, spatial patterns, correlations between weather and animal density, non-synchronized diving), but not mis-identification of duplicate errors – scenarios sc39, 41,43,45,48,50, 52, 54 – the patterns of percent bias of the group size, school and whale density were similar to above, except the magnitude of bias and variability were less than that found in all of the scenarios. Specifically, the estimates of whale density from the SPL method were less bias (average percent bias=0.1, median=2.0, RMS=6.9) then the OK method (average percent bias=6.8, median=10.8, RMS=11.1); and both the OK and SPL were less bias and variable than the standard method (average percent bias=-10.9, median=-11.8, RMS=17.0).

When examining the results of the regressions that investigate how each method dealt with the different factors complicating the analysis of the IDCR/SOWER minke whale data:

<u>OK method</u>

In most scenarios, the OK-2009 method was able to account for the level of biases in the simulations; that is, result in an insignificant factorial effort. However, there were a couple factors that were significant, meaning the OK analysis method was not able to account for the level of the bias that was in the simulated data.

Significant factors include:

9. SPATIAL GRADIENTS. There was a significant negative bias in the density of schools and of whales when there was an interaction between weather conditions and a whale density/school size gradient, also called location correlation. This level of bias was less than that resulting from using the standard estimation method, but more than that resulting from using the SPLINTR method (were the level of bias was not significant).

10. However there not a non-significant bias when there was a vertical (perpendicular to the ice edge) and horizontal (along the ice edge) whale density gradient, only a horizontal whale density gradientor the groups of whales were clustered.

11. NON-SYNCHRONIZED DIVING. There was a significant bias in the density of schools, whales and group size when the whales within a school were surfacing in a non-synchronized manner. However, the direction of the bias was different in the 2005 versus 2008 set of scenarios. The reason for this different is not known.

12. IO/CLOSING MODE. Alternated between IO mode and Closing mode lead to non-significantly biased density estimates of whales and schools, but to a significantly biased estimate of group sizes. However, the direction of the significant bias in group size estimates differed in the 2004 and 2005 sets of scenarios.

13. DUPLICATES. Mis-identifying duplicate sightings lead to very large significant positively biased density estimates of whales, schools and group sizes. This is to be expected by any method.

Non-significant factors include:

14. COVARIATES IN DETECTION FUNCTION. A detection function dependent on weather conditions, school size and on sighting cues (blow versus body, a factor not included in the standard and OK analysis methods) lead to a non-significant bias in the estimated density of whales, schools and group size. However, there were bias estimates when the detection function was dependent on weather conditions and there was a density gradient (see spatial gradients above) or dependent on school size and non-synchronized diving (see non-synchronized diving above).

15. MEASUREMENT ERRORS. Measurement errors in the recorded school size or measurement errors in the recorded radial distances, angles, time and school size resulted in non-significantly biased whale density and average group size estimates.

SPLINTR method

In most scenarios, the SPLINTR-2009 (SPL-2009) method was able to account for the level of biases in the simulations; that is, result in an insignificant factorial effort. However, there were a couple factors that were significant, meaning the SPLINTR analysis method was not able to account for the level of the bias that was in the simulated data.

Significant factors include:

16. NON-SYNCHRONIZED DIVING. There was a significant positive bias in the density of whales and group size when the whales within a school were surfacing in a non-synchronized manner in the 2005 scenarios, but the bias was not significant in the 2008 scenarios. The reason for this different is not known.

17. MEASUREMENT ERRORS. There was a significant positive bias in the density of schools and whales (not group size) when there were measurement errors in the recorded radial distances, angles, time and school size. However if there was only school size measurement error, only the group size estimate was significantly positively biased.

18. COVARIATES IN DETECTION FUNCTION. There was a significant bias in the density of whales, schools and group size when the detection function was dependent on school size, weather or cue in the 2004 scenarios. However, there was a non-significant bias in the 2005 and 2008 scenarios when the detection function was dependent on school size and weather.

19. DUPLICATES. Mis-identifying duplicate sightings lead to very large significant positively biased density estimates of whales, schools and group sizes. This is to be expected by any method.

Non-significant factors include:

20. SPATIAL GRADIENTS. There was no significant bias when there was a whale density gradient, with or without clustering, group size density gradient or density gradients correlated with weather conditions.

21. IO/CLOSING MODE. There was no significant bias when the survey protocols alternated between IO and Closing mode.

Table 1. Factors incorporated into each scenario. Absent indicates that factor is not present, while present indicates that factor is present. DF means detection function. For example, "present" in the DF-cue column indicates the probability of detecting a group is a function of cue. The parameter values of the detection functions are in Table 2.

Table 1A. 2004 set of scenarios

Scenario	Factors						
	Dens Gradient	Size Gradient	Groups Clustered	Mode	DF-size	DF-weather	DF-Cue
1	Absent	Absent	Random	IO+Close	Present	Present	Absent
2	Present	Absent	Random	IO	Absent	Present	Present
3	Absent	Present	Random	IO	Present	Absent	Present
4	Present	Present	Random	IO+Close	Absent	Absent	Absent
5	Absent	Absent	Clustered	IO+Close	Absent	Absent	Present
6	Present	Absent	Clustered	IO	Present	Absent	Absent
7	Absent	Present	Clustered	IO	Absent	Present	Absent
8	Present	Present	Clustered	IO+Close	Present	Present	Present
9	Present	Present	Clustered	IO	Absent	Absent	Present
10	Absent	Present	Clustered	IO+Close	Present	Absent	Absent
11	Present	Absent	Clustered	IO+Close	Absent	Present	Absent
12	Absent	Absent	Clustered	IO	Present	Present	Present
13	Present	Present	Random	IO	Present	Present	Absent
14	Absent	Present	Random	IO+Close	Absent	Present	Present
15	Present	Absent	Random	IO+Close	Present	Absent	Present
16	Absent	Absent	Random	IO	Absent	Absent	Absent

Table 1B. 2005 set of scenarios

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

Table 1C. 2006/7 set of scenarios.

Scenario	Factors						
	School 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
34	Present	Present	Monotonic	Present	Present	Present	Absent
35	Present	Absent	Absent	Absent	Absent	Absent	Absent
36	Present	Present	Monotonic	Present	Present	Present	Absent
37	Absent	Absent	Non-monotonic	Absent	Absent	Absent	Absent
38	Present	Absent	Non-monotonic	Absent	Absent	Absent	Present

Table 1D. 2008 set of scenarios.

Scenario	Factors				
	Measure Errors	Along Ice Density Gradient	Mis-id Dups	Surfacing Patterns	Location Correlations
39	Absent	Present	Absent	In-syn	Absent
40	Present	Absent	Present	In-syn	Absent
41	Present	Present	Absent	Out-of-syn	Absent
42	Absent	Absent	Present	Out-of-syn	Absent
43	Present	Absent	Absent	In-syn	Present
44	Absent	Present	Present	In-syn	Present
45	Absent	Absent	Absent	Out-of-syn	Present
46	Present	Present	Present	Out-of-syn	Present
47	Present	Absent	Present	Out-of-syn	Present
48	Absent	Present	Absent	Out-of-syn	Present
49	Absent	Absent	Present	In-syn	Present
50	Present	Present	Absent	In-syn	Present
51	Absent	Present	Present	Out-of-syn	Absent
52	Present	Absent	Absent	Out-of-syn	Absent
53	Present	Present	Present	In-syn	Absent
54	Absent	Absent	Absent	In-syn	Absent

Table 2A. The actual values and percent bias of the density of whales, density of schools, and the mean school size, along with some summary statistics of each set of scenarios (the average (AVG), median (Med), and Root mean square (RMS) of the percent bias for each scenario.

2004 scenarios

	Actual			Percent Bias								
				Density of whales			Density of Schools			Avg School Size		
Scenario	Density Whales	Density Schools	School size	OK09	SPLINTR	Standard	OK09	SPLINTR	Standard	OK09	SPLINTR	Standard
sc01	0.0715	0.0292	2.45	2.51	-9.22	-15.23	-0.71	-7.81	-18.74	3.24	-1.47	4.29
sc02	0.0706	0.0288	2.45	4.73	-20.34	-35.43	-0.09	-11.17	-33.51	4.91	-10.08	-2.88
sc03	0.0556	0.0291	1.91	11.64	-11.78	-12.78	1.49	-12.56	-13.77	10.07	0.91	1.22
sc04	0.0636	0.0312	2.04	-6.81	-17.73	-27.59	-9.65	-12.92	-29.39	3.15	-5.51	2.53
sc05	0.0712	0.0291	2.45	7.79	-20.24	-27.07	5.58	-14.58	-25.44	2.15	-6.57	-2.15
sc06	0.0704	0.0288	2.44	12.40	-9.21	-4.67	3.36	-10.61	-5.31	8.77	1.59	0.74
sc07	0.0554	0.0291	1.91	2.89	-21.09	-43.66	-1.28	-12.76	-42.63	4.28	-9.35	-1.66
sc08	0.0633	0.0310	2.04	-2.62	-14.16	-17.15	-8.44	-10.88	-23.44	6.47	-3.50	8.22
sc09	0.0581	0.0288	2.02	9.56	-21.27	-25.60	2.36	-16.35	-23.93	7.08	-5.85	-2.14
sc10	0.0571	0.0291	1.96	10.02	-8.00	1.54	3.24	-8.33	-7.11	6.62	0.37	9.36
sc11	0.0727	0.0298	2.44	-17.08	-21.57	-52.16	-16.81	-11.59	-50.73	-0.38	-11.13	-2.92
sc12	0.0707	0.0289	2.44	7.64	-11.11	-10.59	-0.55	-9.45	-12.11	8.15	-1.81	1.68
sc13	0.0584	0.0288	2.03	5.84	-12.13	-18.60	-1.99	-8.83	-18.90	8.10	-3.43	0.45
sc14	0.0566	0.0291	1.95	10.30	-21.49	-31.42	4.80	-12.76	-33.92	5.22	-9.85	3.80
sc15	0.0773	0.0316	2.45	-4.41	-9.92	-19.77	-6.91	-9.91	-22.82	2.76	0.07	4.00
sc16	0.0713	0.0291	2.45	5.34	-15.60	-20.37	-0.40	-11.88	-18.98	5.77	-4.20	-1.68
AVG 01-16	0.0652	0.0295	2.21	3.73	-15.30	-22.53	-1.63	-11.40	-23.80	5.40	-4.36	1.43
Med 01-16	0.0670	0.0291	2.24	5.59	-14.88	-20.07	-0.47	-11.38	-23.13	5.50	-3.85	0.98
RMS 01-16	0.0656	0.0295	2.23	8.55	16.10	26.22	6.05	11.61	26.58	6.03	5.96	3.93

Table 2B. The actual values and percent bias of the density of whales, density of schools, and the mean school size, along with some summary statistics of each set of scenarios (the average (AVG), median (Med), and Root mean square (RMS) of the percent bias for each scenario.
2005 scenarios

	Actual			Percent Bias								
				Density of whales			Density of Schools			Avg School Size		
Scenario	Density Whales	Density Schools	School size	OK09	SPLINTR	Standard	OK09	SPLINTR	Standard	OK09	SPLINTR	Standard
sc17	0.0707	0.0288	2.45	4.21	-9.57	-19.07	-1.23	-7.91	-21.48	5.61	-1.62	3.13
sc18	0.0704	0.0288	2.45	11.10	-7.08	-9.08	3.51	-9.32	-3.05	7.36	2.51	-6.27
sc19	0.0742	0.0303	2.45	1.50	-6.15	-17.11	-0.95	-6.89	-20.32	2.58	0.91	4.12
sc20	0.0772	0.0315	2.45	11.43	-5.52	0.71	6.14	-6.53	-3.01	5.00	1.11	3.90
sc21	0.0715	0.0291	2.45	11.88	-3.37	-11.85	9.82	-10.99	-12.42	2.17	8.83	0.67
sc22	0.0708	0.0289	2.45	6.09	-6.45	-17.48	3.28	-11.72	-28.55	2.87	6.26	15.55
sc23	0.0713	0.0292	2.45	10.65	-5.33	0.92	9.62	-10.86	-13.98	1.15	6.37	17.46
sc24	0.0738	0.0301	2.45	5.17	1.60	-10.41	3.86	-6.84	-28.11	1.48	9.43	24.69
sc25	0.0770	0.0314	2.45	12.65	-3.17	2.67	12.51	-10.23	-11.13	0.33	8.09	15.64
sc26	0.0739	0.0302	2.45	3.89	-3.69	-11.25	2.40	-9.76	-28.65	1.74	7.19	24.53
sc27	0.0705	0.0288	2.45	15.53	-6.56	-4.00	10.96	-12.23	-11.33	4.29	6.60	8.47
sc28	0.0704	0.0288	2.45	7.10	1.34	-18.90	6.31	-5.94	-28.11	1.02	8.04	12.74
sc29	0.0739	0.0302	2.45	3.96	-7.72	-13.92	1.05	-6.81	-18.49	2.91	-0.86	5.61
sc30	0.0711	0.0291	2.44	7.62	-4.35	-2.89	3.65	-6.98	-5.16	3.91	2.85	2.52
sc31	0.0706	0.0288	2.45	2.47	-7.51	-21.91	-1.13	-8.57	-21.42	3.67	1.35	-0.65
sc32	0.0714	0.0292	2.45	11.42	-6.24	-4.97	2.09	-6.96	-5.89	9.20	0.84	1.03
AVG 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
Med 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

Table 2C. The actual values and percent bias of the density of whales, density of schools, and the mean school size, along with some summary statistics of each set of scenarios (the average (AVG), median (Med), and Root mean square (RMS) of the percent bias for each scenario.

2007 scenarios

	Actual			Percent Bias								
				Density of whales			Density of Schools			Avg School Size		
Scenario	Density Whales	Density Schools	School size	OK09	SPLINTR	Standard	OK09	SPLINTR	Standard	OK09	SPL	Standard
sc33	0.0356	0.0145	2.45	4.90	-2.80	-18.40	2.88	-5.86		9.20	3.36	
sc34	0.0295	0.0145	2.03	-0.08	-0.85	-30.00	-4.34	-10.08		2.03	10.85	
sc35	0.1066	0.0436	2.45	7.81	-0.98	-23.04	3.69	-6.40		4.83	5.81	
sc36	0.0921	0.0454	2.03	2.67	-2.40	-31.62	-4.02	-13.30		4.02	13.06	
sc37	0.0361	0.0146	2.47	-1.03	-15.39	-8.99	-4.71	-15.47		7.15	0.13	3.04
sc38	0.0536	0.0219	2.44	7.37	-2.60	-0.10	2.26	-8.96		3.91	7.05	4.00
AVG 33-38	0.0589	0.0258	2.31	3.61	-4.17	-18.69	-0.71	-10.01		5.04	6.71	3.52
Med 33-38	0.0448	0.0183	2.44	3.78	-2.50	-20.72	-0.88	-9.52		4.42	6.43	3.52
RMS 33-38	0.0660	0.0291	2.32	4.96	6.57	21.79	3.74	10.59	0.00	5.69	7.99	2.05

Table 2D. The actual values and percent bias of the density of whales, density of schools, and the mean school size, along with some summary statistics of each set of scenarios (the average (AVG), median (Med), and Root mean square (RMS) of the percent bias for each scenario.
2008 scenarios

	Actual			Percent Bias								
				Density of whales			Density of Schools			Avg School Size		
Scenario	Density Whales	Density Schools	School size	OK09	SPLINTR	Standard	OK09	SPLINTR	Standard	OK09	SPL	Standard
sc39	0.1018	0.0291	3.50	11.62	-1.38	-1.67	3.75	-7.12	-4.52	7.64	6.25	2.91
sc41	0.0715	0.0292	2.45	14.87	7.59	5.23	16.49	-0.95	-5.51	-1.16	9.05	11.63
sc43	0.0762	0.0291	2.62	8.83	3.86	-23.81	-8.12	-2.41	-24.77	18.31	7.01	1.34
sc45	0.0586	0.0292	2.01	-9.40	-12.18	-25.08	-16.53	-19.93	-39.71	8.87	10.25	24.43
sc48	0.0587	0.0291	2.02	-6.73	-9.18	-24.13	-13.51	-15.73	-38.43	8.21	8.80	23.40
sc50	0.0764	0.0291	2.62	9.98	4.90	-21.94	-5.94	-3.07	-23.11	16.90	8.66	1.71
sc52	0.0714	0.0291	2.45	12.98	6.93	5.16	14.66	-1.32	-6.05	-1.19	8.86	12.17
sc54	0.1024	0.0291	3.52	11.98	0.11	-0.93	3.54	-6.37	-3.06	8.12	6.95	1.87
AVG 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
Med 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
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
AVG 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
Med 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
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

Table 2E. The actual values and percent bias of the density of whales, density of schools, and the mean school size, along with some summary statistics of each set of scenarios (the average (AVG), median (Med), and Root mean square (RMS) of the percent bias for each scenario.

2008 scenarios

	Actual			Percent Bias								
				Density of whales			Density of Schools			Avg School Size		
Scenario	Density Whales	Density Schools	School size	OK09	SPLINTR	Standard	OK09	SPLINTR	Standard	OK09	SPL	Standard
Scenarios with mis-identification of duplicates												
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
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
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
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
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
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
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
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
AVG 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
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

* Scenarios with mis-identification of duplicates

** Average and RMS are without the mis-identification of duplicate scenarios

Table 3. Factorial effects of each factor when estimating the **density of individual whales** for each analysis method. That is, the difference between the two treatment levels. DF means detection function. RMSE is the root mean square error of the factorial effects. Avg is the average of the factorial effects. A. Scenarios from 2004. B. Scenarios from 2005. C. Scenarios from 2008. OK is the OK analysis method; SPL is the SPLINTR analysis method; and INT is the Integrated analysis method.

3A. 2004 (sc01-sc16) DENSITY OF WHALES

Analysis Method	Density gradient	Size gradient	Clustered	Mode change	DF with size effect	DF with weather effect	DF with cue effect	RMSE	Avg
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
INT – 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

3B. 2005 (sc17-sc32) DENSITY OF WHALES

Analysis Method	School size error	Weather and density interaction	Vertical and horizontal density	Mode changes	Non-synchronized surfacing	RMSE	Avg
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
INT – 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

3C. 2008 (sc39-sc54) DENSITY OF WHALES

Analysis Method	Measure errors	Along ice density gradient	Mis-id dups	Surfacing patterns	Location correlation	RMSE**	Avg**
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

* Indicates the level of bias is significant at the 0.01 p-value level.
** Does not include the mis-identification duplicates factor.
+ Indicates there is a significant interaction between the factors within the same row.

Table 4. Factorial effects of each factor when estimating the **MEAN GROUP SIZE** for each analysis method. That is, the difference between the two treatment levels. DF means detection function. RMSE is the root mean square error of the factorial effects. Avg is the average of the factorial effects. A. Scenarios from 2004. B. Scenarios from 2005. C. Scenarios from 2008. OK is the OK analysis method; SPL is the SPLINTR analysis method; and INT is the Integrated analysis method.

4A. 2004 (sc01-sc16) MEAN GROUP SIZE

Analysis Method	Density gradient	Size gradient	Clustered	Mode change	DF with size effect	DF with weather effect	DF with cue effect	RMSE	Avg
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
Standard									

4B. 2005 (sc17-sc32) MEAN GROUP SIZE

Analysis Method	School size error	Weather and density interaction	Vertical and horizontal density	Mode changes	Non-synchronized surfacing	RMSE	Avg
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
Standard							

4C. 2008 (sc39-sc54) MEAN GROUP SIZE

Analysis Method	Measure errors	Along ice density gradient	Mis-id dups	Surfacing patterns	Location correlations	RMSE**	Avg**
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
Standard							

* Indicates the level of bias is significant at the 0.01 p-value level.

** Does not include the mis-identification duplicates factor.

+ Indicates there is a significant interaction between the factors within the same row.

Table 5. Factorial effects of each factor when estimating the **DENSITY OF SCHOOLS** for each analysis method. That is, the difference between the two treatment levels. DF means detection function. RMSE is the root mean square error of the factorial effects. Avg is the average of the factorial effects. A. Scenarios from 2004. B. Scenarios from 2005. C. Scenarios from 2008. OK is the OK analysis method; SPL is the SPLINTR analysis method; and INT is the Integrated analysis method.

5A. 2004 (sc01-sc16) DENSITY OF SCHOOLS

Analysis Method	Density gradient	Size gradient	Clustered	Mode change	DF with size effect	DF with weather effect	DF with cue effect	RMSE	Avg
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
Standard									

5B. 2005 (sc17-sc32) DENSITY OF SCHOOLS

Analysis Method	School size error	Weather and density interaction	Vertical and horizontal density	Mode changes	Non-synchronized surfacing	RMSE	Avg
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
Standard							

5C. 2008 (sc39-sc54) DENSITY OF SCHOOLS

Analysis Method	Measure errors	Along ice density gradient	Mis-id dups	Surfacing patterns	Location correlation	RMSE**	Avg**
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
Standard							

* Indicates the level of bias is significant at the 0.01 p-value level.
** Dos not include the mis-identification duplicates factor.
+ Indicates there is a significant interaction between the factors within the same row.

Figure 1. Box plot of the percent bias ((estimated-actual)/actual) of the DENSITY OF WHALES each of the100 iterations within a scenario. Scenarios with X are those with mis-identified duplicate errors.

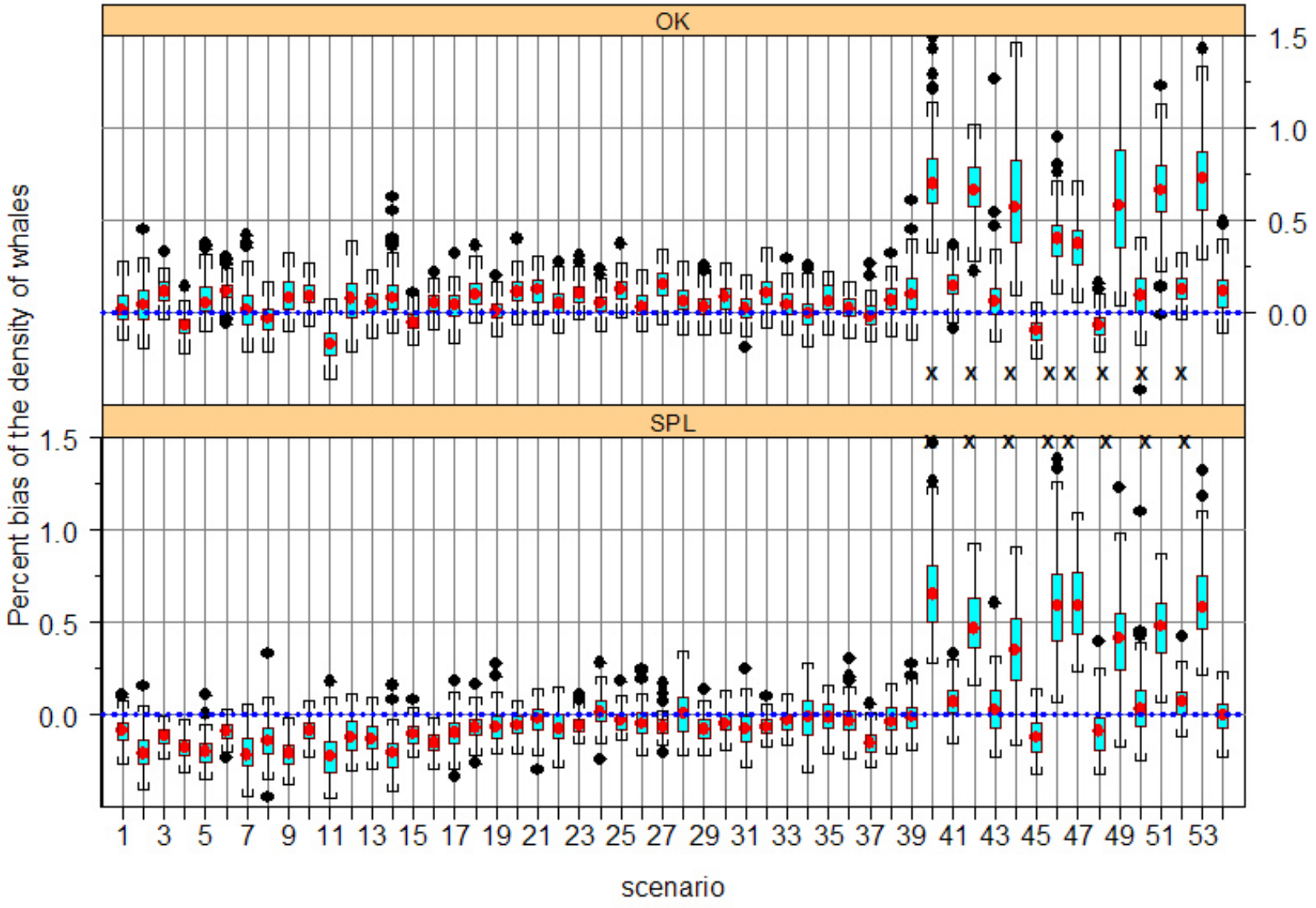

Figure 2. Box plot of the percent bias ((estimated-actual)/actual) of the DENSITY OF SCHOOLS each of the100 iterations within a scenario. Scenarios with X are those with mis-identified duplicate errors.

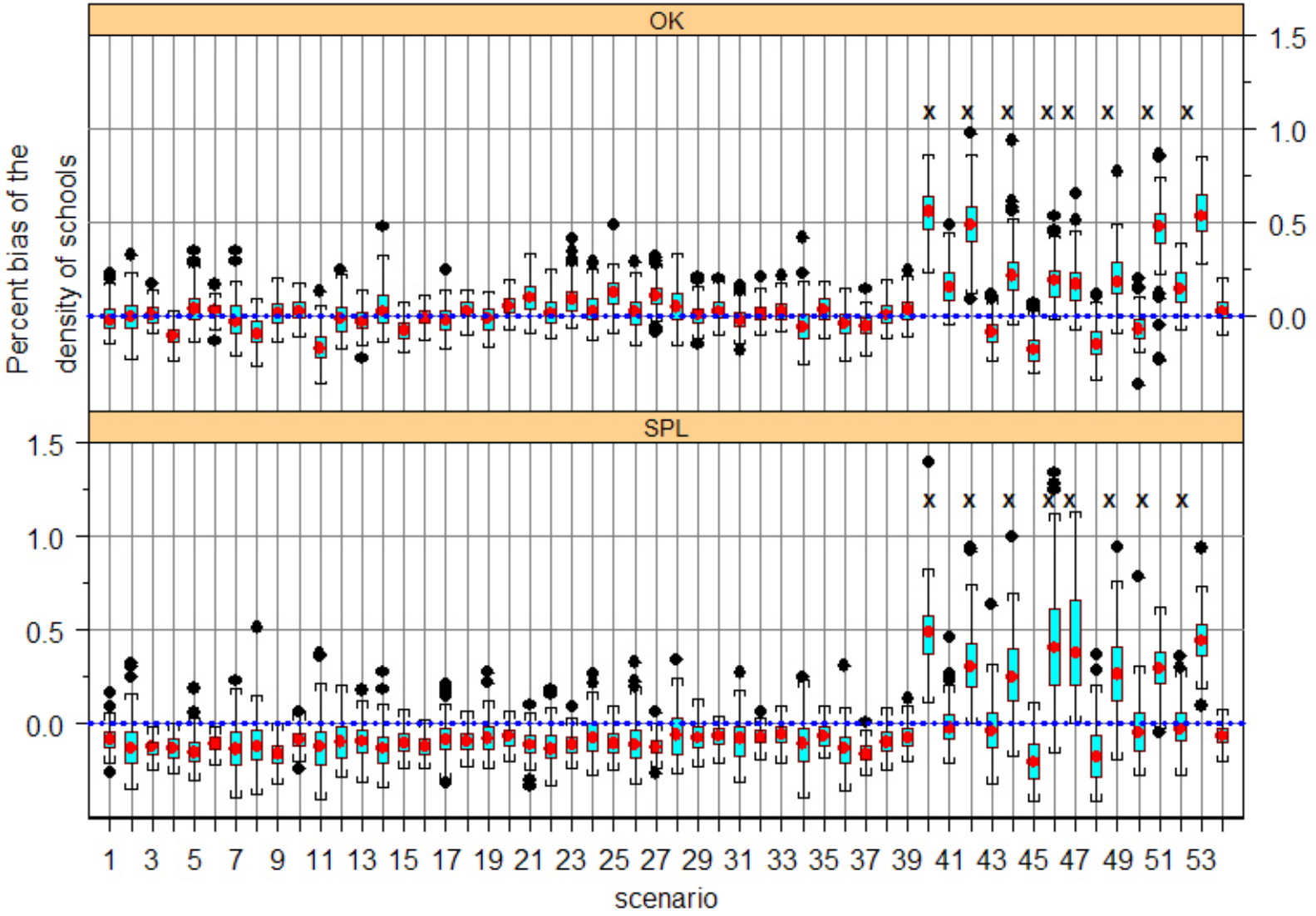

Figure 3. Box plot of the percent bias ((estimated-actual)/actual) of the MEAN GROUP SIZES each of the100 iterations within a scenario. Scenarios with X are those with mis-identified duplicate errors.

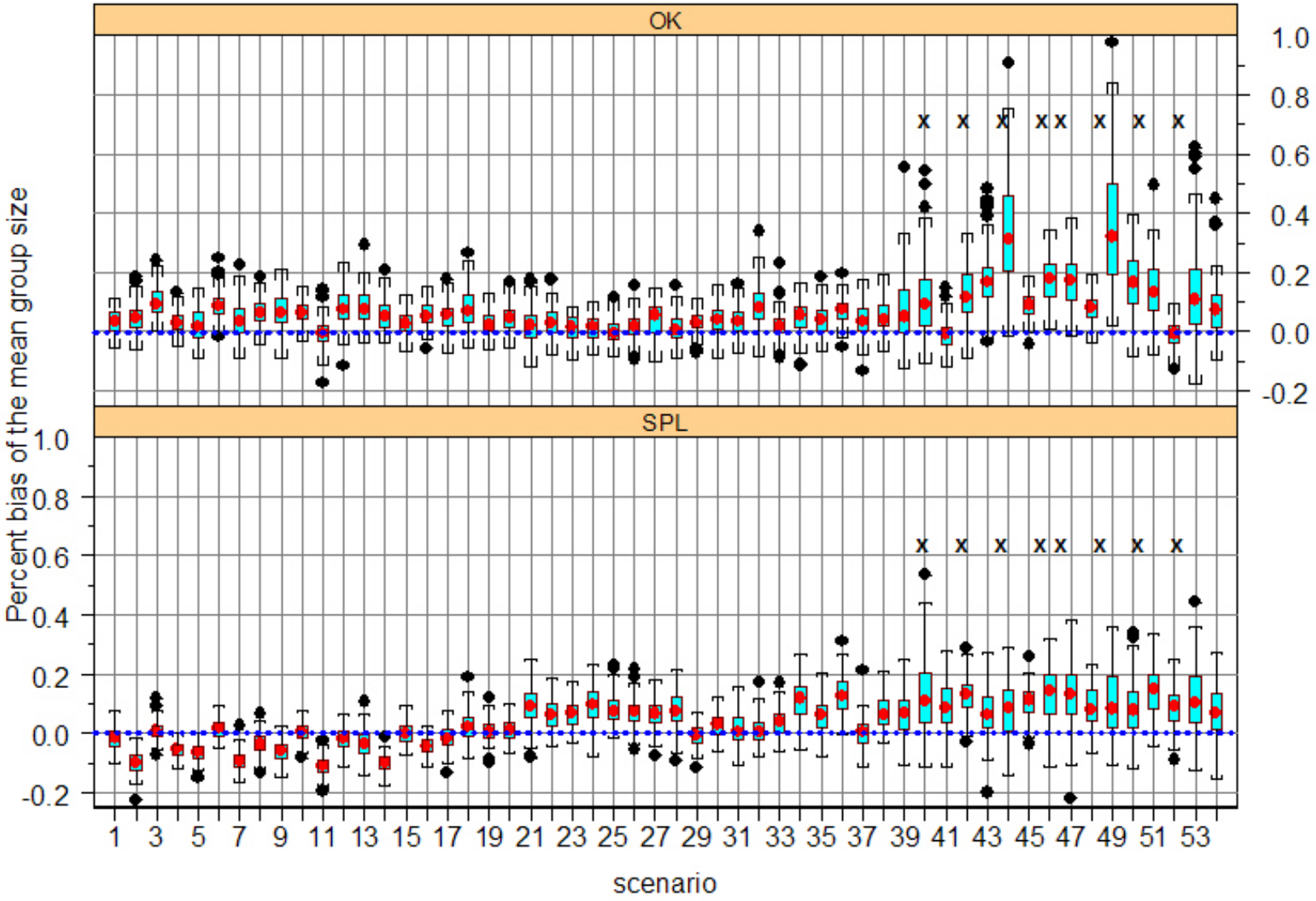

Figure 4. Distribution of the percent differences of each iteration within a scenario when comparing the results for the OK versus the SPLINTR methods.

FIRST SET OF SCENARIOS – 2004

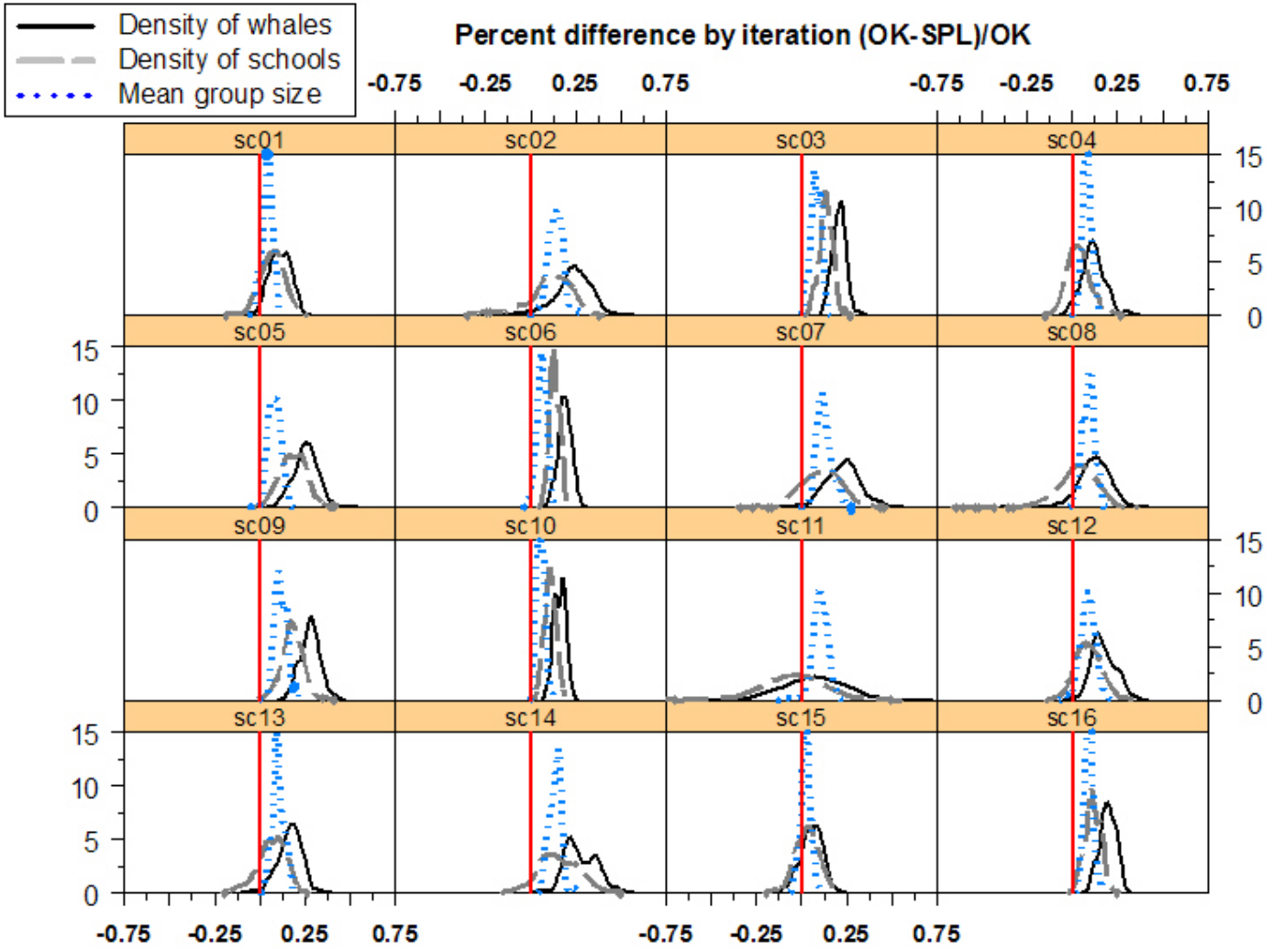

Figure 5. Distribution of the percent differences of each iteration within a scenario when comparing the results for the OK versus the SPLINTR methods.

SECOND SET OF SCENARIOS – 2005

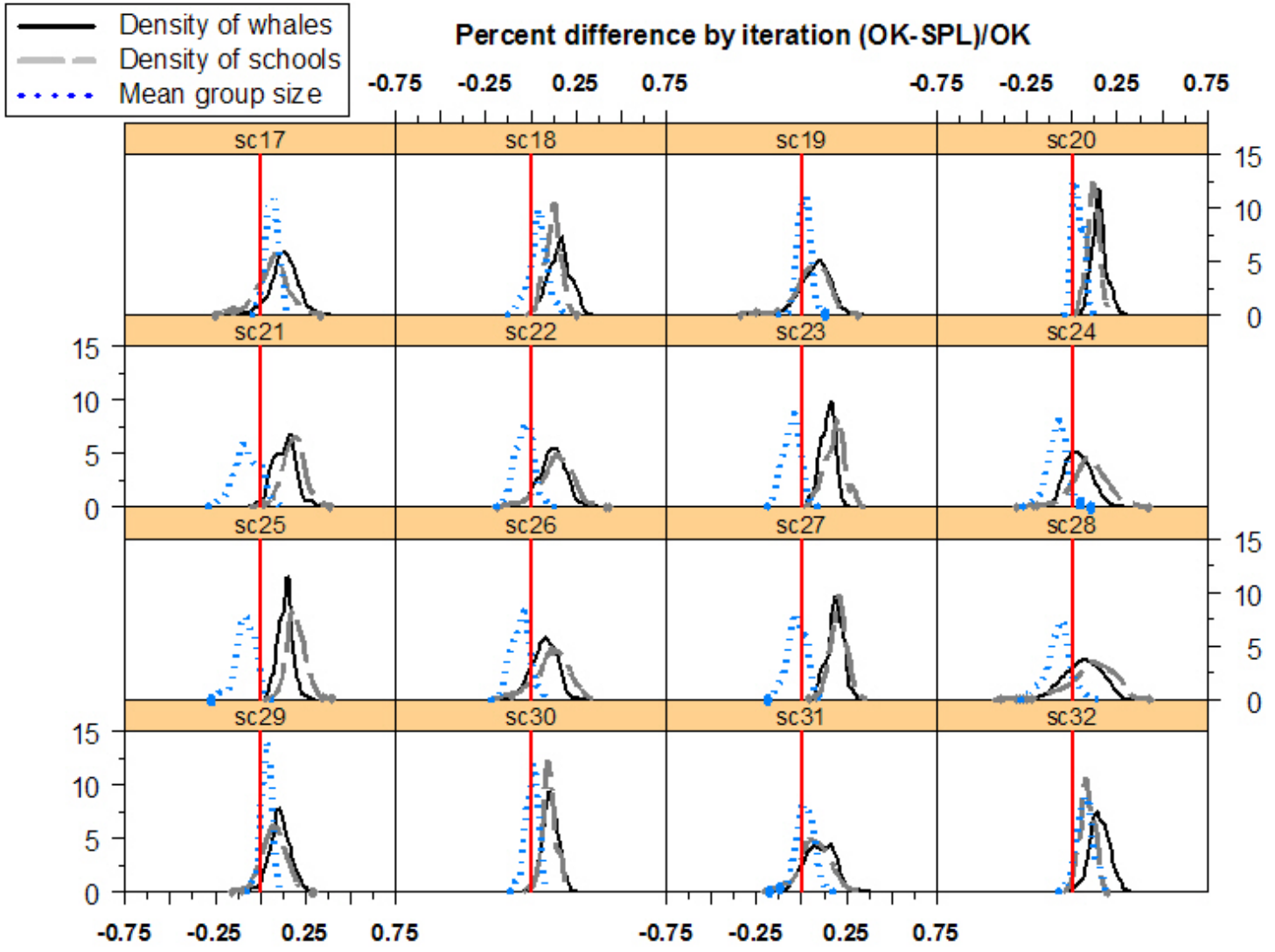

Figure 6. Distribution of the percent differences of each iteration within a scenario when comparing the results for the OK versus the SPLINTR methods.

THIRD SET OF SCENARIOS – 2008