Applying the direct duplicate method to simulated IDCR/SOWER survey data

T.A. BRANCH[#] AND D.S. BUTTERWORTH[#]

ABSTRACT

The direct duplicate method (Palka, 1994) was applied to simulated IDCR/SOWER survey data. Estimates of whale density were generally negative biased, but less so than estimates obtained using the standard method. The mean bias across scenarios was -11% (range -31% to 8%) for the "2004" scenarios and -5% (range -19% to 10%) for the "2005" scenarios. Negative bias was more pronounced when a density gradient was present, when the detection function used to generate the simulated sightings excluded school size but included weather as a covariate, when errors in recorded school size were introduced, when weather and density were correlated, and when surveys were conducted in IO mode only. This method shows promise although further development is desirable to reduce the associated bias further, perhaps by including weather and school size as covariates.

INTRODUCTION

Estimates of circumpolar abundance for Antarctic minke whales (*Balaenoptera bonaerensis*) have traditionally been obtained using the "standard" method of applying line transect methodology to the IDCR/SOWER sightings data (e.g. Branch and Butterworth, 2001). The standard method assumes that $g(0)=1$, where $g(y)$ is the probability of detecting a school at perpendicular distance y from the trackline, implying that no schools on the trackline are missed. Since some schools on the trackline undoubtedly are missed, estimates using the standard method will likely be negatively biased.

Three alternative methods (Cooke, 2001; Bravington, 2004; Okamura and Kitakado, 2004) attempt to address this problem by estimating $g(0)$ from the Independent Observer (IO) data in the IDCR/SOWER surveys. In IO mode, sightings are recorded from two independent platforms. To test the efficacy of these new analytical methods, simulated IDCR/SOWER-like data were generated by Palka and Smith (2004; 2005). Analyses of these data using the standard method confirmed that the standard method produced negatively biased estimates of whale density (Branch 2005). There are several variants of line transect methodology that can be used to estimate whale density without assuming that $g(0)=1$. In this paper we apply the "direct duplicate" method (Palka 1995) to the simulated data (Palka and Smith, 2004; 2005).

METHODS

Simulated survey data

Input data files were generated as described in Palka and Smith (2004; 2005). Each survey iteration was conducted for 30-120 days depending on the scenario. The survey vessel travelled perpendicular to the ice edge for 16 hours and then shifted 20 km to the east for the following day's survey. Half of the scenarios were conducted in independent observer (IO) mode, the other half alternated between IO mode and closing mode. Scenarios included various covariates or assumptions and were divided into the "2004" scenarios (sc01 to sc16) and the "2005" scenarios (sc17 to sc32). The simulations were never intended to mimic exactly the IDCR/SOWER survey conditions but instead to test estimation methods, therefore estimates of $g(0)$ obtained from the simulated data are not indicative of the true $g(0)$ encountered on the surveys. Some differences between the simulations and the IDCR/SOWER surveys were summarised in Branch (2005).

Direct duplicate method

The direct duplicate method uses standard distance sampling methodology to provide separate estimates of density from sightings made from the IO platform, the barrel (topman position), and those made from both platforms (duplicate sightings). The method then uses the Petersen two-sample mark-

[#] *MARAM (Marine Research Assessment and Management Group), Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch, 7701, South Africa.e-mail: tbranch@maths.uct.ac.za*

recapture equation to obtain the density of whales by treating the duplicates as "recaptures" of animals that were marked by one of the platforms. The overall estimated density of whales (D_w) is obtained from:

$$D_w = \frac{D_{w,IO} D_{w,topman}}{D_{w,duplicates}} \qquad (1)$$

where:

$D_{w,IO}$ = density calculated from all sightings seen from the IO platform

$D_{w,topman}$ = density calculated from all sightings seen from the topman position

$D_{w,duplicates}$ = density calculated from all sightings that were seen by both the IO and topman positions.

Mark-recapture methods of estimating $g(0)$ make the assumption of independence of sightings by different observers, so that bias is introduced when heterogeneity in sightability of observer efficiency results in violation of this assumption. The attractiveness of the direct duplicate method of $g(0)$ estimation is that it requires the lesser assumption that independence applies only on the trackline, so that any bias resulting from non-independence is likely to be reduced (Buckland *et al*. 2004, p.133).

Two variants of this equation were needed, one for the scenarios that contained IO mode only data, and another for the scenarios consisting of alternating IO and closing mode data. For scenarios which consisted only of IO mode data, all sightings from the bridge were excluded and then the methods of Branch (2005) were applied to obtain the components of equation (1). Perpendicular distances had to be recalculated for duplicate and triplicate sightings because the perpendicular distance of the first such sighting had been recorded for all associated sightings in the simulated sightings files, but the direct duplicate method required platform-specific perpendicular distances.

For scenarios that consisted of a combination of IO and closing mode, the method above was used to obtain the estimated density of schools (D_s) from the IO mode portion of the surveys. The method could not be applied to the closing mode portion of the surveys since duplicate sightings were not recorded in closing mode. However, estimated school size, $E[s]_{\text{closing}}$, was obtained using closing mode sightings from all platforms, and then the estimated density of whales, D_w, was obtained as follows:

$$D_s = \frac{D_{s,\text{IO}} D_{s,\text{topman}}}{D_{s,\text{duplicates}}} \qquad (2)$$
$$D_w = D_s \cdot E[s]_{\text{closing}}$$

The conversion factor between IO and closing mode were not considered, nor did we calculate the inverse-variance weighted estimates of density obtained by combining IO mode estimates and closing mode estimates.

RESULTS

Results for "2004" scenarios

Estimated effective search half width was much greater for the topman platform (mean across scenarios 1467 m, range 1112–1675 m, Table 1) than for the IO platform (mean 1149 m, range 823–1317 m), and was smallest for duplicate sightings (mean 895 m, range 569–1104 m). There was a strong correlation (across scenarios) between effective search half width recorded from the IO platform and the topman platform (r = 0.97), between the IO platform and duplicate sightings (r = 0.97) and between the topman and duplicate sightings (r = 0.91).

Estimated mean school size was generally close to the true values (Table 2). In scenarios with only IO mode, there was a tendency for mean school size to be greater when estimated effective search half width was smaller, thus the largest mean school size estimates were obtained for duplicate sightings, and the smallest for topman sightings.

The estimated density of schools (and whales) was greatest when based on IO platform sightings, and least when based on duplicate sightings (Tables 3, 4).

There was a negative bias in whale density estimated by the direct duplicate method (mean -11%, range -31% to 8%, Table 4), but the estimates from individual iterations displayed a wide spread of values that generally spanned the true values (Figure 1). Estimated densities displayed greater negative bias when a density gradient was present (-14% vs. -8%), when the true detection function (i.e. that used to generate the simulated sightings) did not include school size as a covariate (-15% vs. -7%), and when the true detection function included weather as a covariate (-16% vs. -6%), but other factors had little influence on bias. The implied $g(0)$ had a mean of 0.89 across scenarios (range 0.70 to 1.08).

Results for "2005" scenarios

Similar patterns were obtained as for the "2004" scenarios: greatest effective search half width, smallest mean school size and highest density of schools was obtained from the topman platform, and the smallest search half width, largest mean school size and lowest density of schools was obtained from duplicate sightings, with estimates from the IO platform intermediate.

The estimated whale density from the direct duplicate method was negatively biased for 12 of 16 scenarios (mean bias -5%, range -19 to 10%, Table 4, Figure 2). Estimated density was more biased when school size errors were present (-7% vs. -4%), when density and weather were correlated (-12% vs. 1%), and when only IO mode survey was conducted (-11% vs. 0%), but the other factors had only a negligible impact on the bias associated with the estimated density of whales. The implied $g(0)$ had a mean over scenarios of 0.95 (range 0.81 to 1.10).

DISCUSSION

Estimated whale densities from the direct duplicate method were generally negatively biased (-11% for "2004" scenarios, -5% for "2005" scenarios) but less so than estimates from the standard method (-23% and -10% respectively (Branch 2005)). Compared to the standard method, similar factors caused negative bias: introducing school size and weather as covariates in the detection function, introducing errors in school size estimation, correlations between density and weather, and when surveys were conducted in IO mode only. It is interesting that estimates of density from simulated surveys in IO mode were negatively biased whereas simulations that also included closing mode were essentially unbiased. This effect likely resulted from negative bias in school size estimates during IO mode.

In summary, applying the direct duplicate method to IDCR/SOWER data would likely eliminate some of the negative bias from assuming that $g(0)=1$, especially if closing mode estimates of school size were used. Further reduction in bias may be possible if covariates such as weather and school size were incorporated.

ACKNOWLEDGEMENTS

Jeff Laake suggested applying the direct duplicate method to the simulated data. The authors are grateful for financial assistance from the International Whaling Commission and the South African National Antarctic Programme.

REFERENCES

Branch, T.A. 2005. Estimated density of Antarctic minke whales obtained from simulated IDCR/SOWER survey data using the "standard method". Paper SC/57/IA14 presented to the IWC Scientific Committee, June 2005 (unpublished). 16pp.

Branch, T.A. and 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*. 3:143-174.

Bravington, M. 2004. A new spatial abundance estimator, applied to simulated survey data. Paper SC/56/IA3 presented to the IWC Scientific Committee, July 2004 (unpublished). 12pp.

Buckland, S.T., Anderson. D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. and Thomas, L. 2004. *Advanced Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press, Oxford. xvii+416pp.

Cooke, J.G. 2001. A modification of the radial distance method for dual-platform line transect analysis, to improve robustness. Paper SC/53/IA31 presented to the IWC Scientific Committee, July 2001 (unpublished). 7pp.

Okamura, H. and Kitakado, T. 2004. Advance in an abundance estimation model of Antarctic minke whales. Paper SC/56/IA9 presented to the IWC Scientific Committee, July 2004 (unpublished). 12pp.

Palka, D. 1995. Abundance estimate of the Gulf of Maine harbor porpoise. *Rep. int. Whal. Commn (Special Issue)* 16:27-50.

Palka, D.L. and Smith, D.W. 2004. Simulating the IDCR/SOWER surveys - 2004. Paper SC/56/IA6 presented to the IWC Scientific Committee, July 2004 (unpublished). 16pp.

Palka, D.L. and Smith, D.W. 2005. Description of 2005 simulations of the IWC/SOWER Southern Hemisphere minke whale abundance surveys. Paper SC/57/IA2 presented to the IWC Scientific Committee, June 2005 (unpublished). 8pp.

Table 1. Estimated effective search half-width (m) and associated CV for each scenario based on sightings in IO mode from the IO platform, from the topman platform and from duplicate sightings made from both the IO and topman platforms.

	IO platform		Topman platform		Duplicate sightings	
Scenario	esw	CV	esw	CV	esw	CV
sc01	1087	0.099	1377	0.086	899	0.101
sc02	1179	0.132	1487	0.073	909	0.188
sc03	1201	0.094	1543	0.074	912	0.111
sc04	1178	0.099	1536	0.081	893	0.139
sc05	1144	0.117	1459	0.079	849	0.179
sc06	1317	0.066	1675	0.068	1056	0.081
sc07	823	0.101	1112	0.105	569	0.175
sc08	1317	0.131	1565	0.109	1104	0.162
sc09	1127	0.109	1462	0.082	833	0.153
sc10	1246	0.086	1638	0.086	995	0.106
sc11	839	0.125	1120	0.088	591	0.191
sc12	1284	0.114	1554	0.096	1089	0.131
sc13	1076	0.121	1373	0.093	875	0.139
sc14	1149	0.122	1462	0.088	899	0.197
sc15	1229	0.099	1579	0.078	959	0.136
sc16	1180	0.086	1524	0.067	886	0.112
sc17	910	0.113	1231	0.092	732	0.150
sc18	1293	0.100	1668	0.089	1042	0.117
sc19	898	0.111	1215	0.087	718	0.133
sc20	1299	0.102	1661	0.090	1040	0.125
sc21	1323	0.101	1675	0.083	1135	0.111
sc22	916	0.106	1229	0.083	773	0.125
sc23	1308	0.100	1682	0.084	1105	0.121
sc24	924	0.110	1254	0.091	775	0.140
sc25	1330	0.098	1674	0.102	1133	0.119
sc26	929	0.115	1244	0.095	791	0.147
sc27	1324	0.120	1654	0.091	1110	0.130
sc28	914	0.129	1228	0.096	774	0.154
sc29	897	0.101	1202	0.090	724	0.125
sc30	1295	0.103	1672	0.077	1035	0.109
sc31	918	0.106	1240	0.093	737	0.151
sc32	1313	0.091	1665	0.073	1068	0.106

Table 2. Estimated school size and associated CV for each of the scenarios. For scenarios which only contained IO mode data, estimated school size was calculated separately from sightings made from the IO platform, those made from the topman platform, and those made from both the IO and topman platforms. For scenarios which included closing mode data, the estimated school size was calculated from sightings made from all platforms during closing mode.

	IO platform		Topman platform		Duplicate sightings		All platforms		
Scenario	E[s]	CV	E[s]	CV	E[s]	CV	E[s]	CV	True E[s]
sc01	–	–	–	–	–	–	2.56	0.072	2.45
sc02	2.35	0.064	2.37	0.055	2.33	0.095	–	–	2.45
sc03	2.04	0.064	1.97	0.052	2.14	0.077	–	–	1.91
sc04	–	–	–	–	–	–	2.09	0.059	2.04
sc05	–	–	–	–	–	–	2.39	0.070	2.45
sc06	2.58	0.043	2.49	0.039	2.67	0.050	–	–	2.44
sc07	1.87	0.073	1.89	0.064	1.86	0.112	–	–	1.91
sc08	–	–	–	–	–	–	2.21	0.078	2.04
sc09	1.99	0.072	1.97	0.051	1.99	0.085	–	–	2.02
sc10	–	–	–	–	–	–	2.15	0.071	1.96
sc11	–	–	–	–	–	–	2.37	0.054	2.44
sc12	2.57	0.069	2.53	0.063	2.66	0.082	–	–	2.44
sc13	2.12	0.065	2.05	0.059	2.19	0.084	–	–	2.03
sc14	–	–	–	–	–	–	2.02	0.075	1.95
sc15	–	–	–	–	–	–	2.54	0.049	2.45
sc16	2.41	0.056	2.41	0.046	2.41	0.073	–	–	2.45
sc17	2.64	0.068	2.58	0.054	2.76	0.086	–	–	2.45
sc18	2.49	0.061	2.38	0.063	2.61	0.068	–	–	2.45
sc19	–	–	–	–	–	–	2.55	0.056	1.91
sc20	–	–	–	–	–	–	2.55	0.083	2.04
sc21	2.90	0.056	2.72	0.057	3.29	0.063	–	–	2.45
sc22	3.26	0.068	3.08	0.053	3.79	0.087	–	–	2.44
sc23	–	–	–	–	–	–	2.87	0.069	1.91
sc24	–	–	–	–	–	–	3.06	0.055	2.04
sc25	–	–	–	–	–	–	2.83	0.079	2.02
sc26	–	–	–	–	–	–	3.05	0.067	1.96
sc27	3.07	0.063	2.88	0.059	3.41	0.079	–	–	2.44
sc28	3.19	0.061	2.99	0.051	3.76	0.075	–	–	2.44
sc29	–	–	–	–	–	–	2.58	0.065	2.03
sc30	–	–	–	–	–	–	2.50	0.083	1.95
sc31	2.56	0.058	2.47	0.048	2.67	0.070	–	–	2.45
sc32	2.59	0.060	2.51	0.058	2.68	0.074	–	–	2.45

Table 3. Estimated density of schools based on IO platform sightings, topman platform sightings, and duplicate sightings from both the IO and topman platforms.

	IO platform		Topman platform		Duplicate sightings	
Scenario	Ds	CV	Ds	CV	Ds	CV
sc01	0.0155	0.125	0.0191	0.099	0.0116	0.138
sc02	0.0104	0.144	0.0132	0.083	0.0060	0.201
sc03	0.0144	0.104	0.0190	0.077	0.0100	0.138
sc04	0.0135	0.121	0.0165	0.089	0.0085	0.158
sc05	0.0112	0.140	0.0162	0.100	0.0067	0.203
sc06	0.0202	0.086	0.0235	0.078	0.0167	0.105
sc07	0.0086	0.111	0.0107	0.113	0.0040	0.215
sc08	0.0170	0.166	0.0194	0.137	0.0132	0.201
sc09	0.0115	0.127	0.0161	0.086	0.0069	0.166
sc10	0.0200	0.113	0.0232	0.088	0.0161	0.130
sc11	0.0076	0.150	0.0094	0.099	0.0034	0.225
sc12	0.0184	0.153	0.0205	0.127	0.0142	0.175
sc13	0.0152	0.121	0.0187	0.098	0.0114	0.157
sc14	0.0104	0.132	0.0132	0.100	0.0058	0.212
sc15	0.0145	0.108	0.0188	0.081	0.0100	0.157
sc16	0.0141	0.090	0.0176	0.075	0.0091	0.120
sc17	0.0142	0.123	0.0174	0.099	0.0103	0.157
sc18	0.0205	0.129	0.0238	0.091	0.0170	0.149
sc19	0.0155	0.126	0.0188	0.092	0.0112	0.148
sc20	0.0227	0.122	0.0266	0.101	0.0188	0.151
sc21	0.0177	0.116	0.0208	0.093	0.0137	0.138
sc22	0.0128	0.120	0.0157	0.095	0.0090	0.155
sc23	0.0178	0.113	0.0208	0.090	0.0138	0.126
sc24	0.0137	0.121	0.0164	0.107	0.0097	0.167
sc25	0.0195	0.127	0.0233	0.110	0.0153	0.146
sc26	0.0134	0.126	0.0163	0.108	0.0093	0.174
sc27	0.0179	0.130	0.0212	0.101	0.0142	0.140
sc28	0.0129	0.149	0.0156	0.113	0.0091	0.181
sc29	0.0154	0.113	0.0191	0.092	0.0111	0.148
sc30	0.0203	0.115	0.0237	0.095	0.0166	0.126
sc31	0.0143	0.117	0.0174	0.103	0.0104	0.152
sc32	0.0203	0.102	0.0238	0.076	0.0166	0.122

Table 4. Estimated density of whales obtained from the IO platform, the topman platform, duplicates, and using the direct duplicate (DD) method. The percent bias = $100(\text{observed} - \text{actual})/\text{actual}$ and the implied $g(0)$ are also shown.

Scenario	IO	Topman	Duplicates	DD	DD CV	True D_w	Bias (%)	$g(0)$
sc01	0.040	0.049	0.030	0.066	0.144	0.072	-8.3	0.92
sc02	0.024	0.031	0.014	0.056	0.151	0.071	-20.7	0.79
sc03	0.029	0.038	0.021	0.052	0.098	0.056	-6.7	0.93
sc04	0.028	0.035	0.018	0.055	0.135	0.064	-13.3	0.87
sc05	0.027	0.039	0.016	0.066	0.157	0.071	-7.7	0.92
sc06	0.052	0.059	0.045	0.069	0.083	0.070	-2.2	0.98
sc07	0.016	0.020	0.007	0.045	0.187	0.055	-19.3	0.81
sc08	0.038	0.043	0.029	0.056	0.160	0.063	-12.2	0.88
sc09	0.023	0.032	0.014	0.054	0.149	0.058	-7.7	0.92
sc10	0.043	0.050	0.035	0.062	0.127	0.057	8.3	1.08
sc11	0.018	0.022	0.008	0.051	0.177	0.073	-30.5	0.69
sc12	0.047	0.052	0.038	0.065	0.146	0.071	-8.0	0.92
sc13	0.032	0.038	0.025	0.050	0.110	0.058	-14.8	0.85
sc14	0.021	0.027	0.012	0.049	0.139	0.057	-13.7	0.86
sc15	0.037	0.048	0.025	0.070	0.114	0.077	-9.8	0.90
sc16	0.034	0.042	0.022	0.066	0.106	0.071	-7.5	0.93
sc17	0.037	0.045	0.028	0.060	0.127	0.071	-15.4	0.85
sc18	0.051	0.057	0.044	0.065	0.123	0.070	-7.2	0.93
sc19	0.040	0.048	0.029	0.066	0.124	0.074	-10.6	0.89
sc20	0.058	0.068	0.048	0.082	0.123	0.077	6.4	1.06
sc21	0.051	0.057	0.045	0.065	0.113	0.072	-9.8	0.90
sc22	0.042	0.048	0.034	0.059	0.114	0.071	-16.0	0.84
sc23	0.051	0.060	0.040	0.077	0.143	0.071	8.3	1.08
sc24	0.042	0.050	0.030	0.071	0.134	0.074	-3.6	0.96
sc25	0.055	0.066	0.043	0.085	0.136	0.077	10.0	1.10
sc26	0.041	0.050	0.028	0.072	0.118	0.074	-3.0	0.97
sc27	0.055	0.061	0.048	0.069	0.106	0.071	-1.9	0.98
sc28	0.041	0.047	0.034	0.057	0.136	0.070	-19.0	0.81
sc29	0.040	0.049	0.029	0.069	0.129	0.074	-6.3	0.94
sc30	0.051	0.059	0.042	0.073	0.115	0.071	2.0	1.02
sc31	0.037	0.043	0.028	0.057	0.141	0.071	-18.9	0.81
sc32	0.053	0.060	0.044	0.071	0.091	0.071	-0.7	0.99

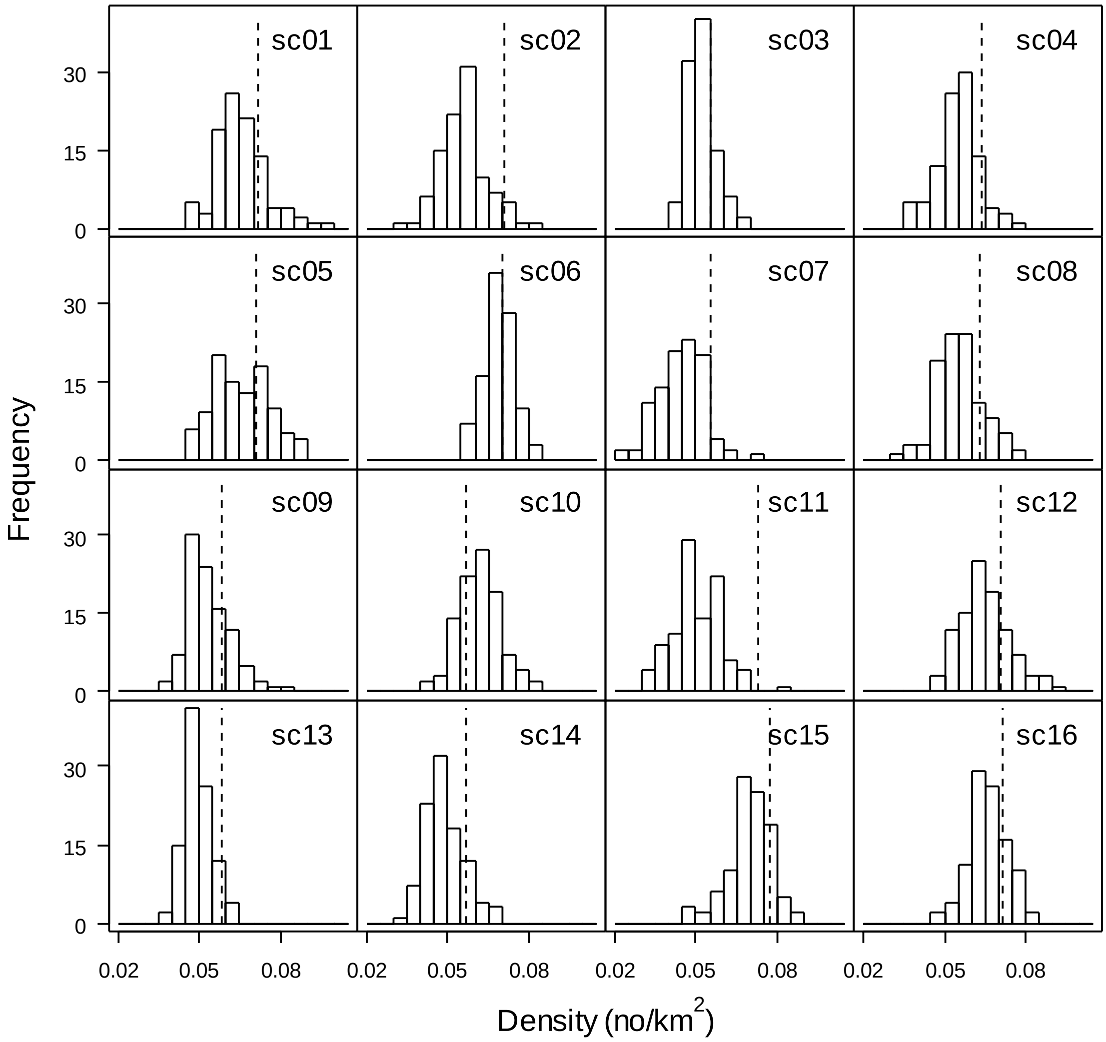

Figure 1. Histograms of estimated density (whales per km^2) for each of the "2004" scenarios. The true mean values in the simulations are indicated by dashed vertical lines.

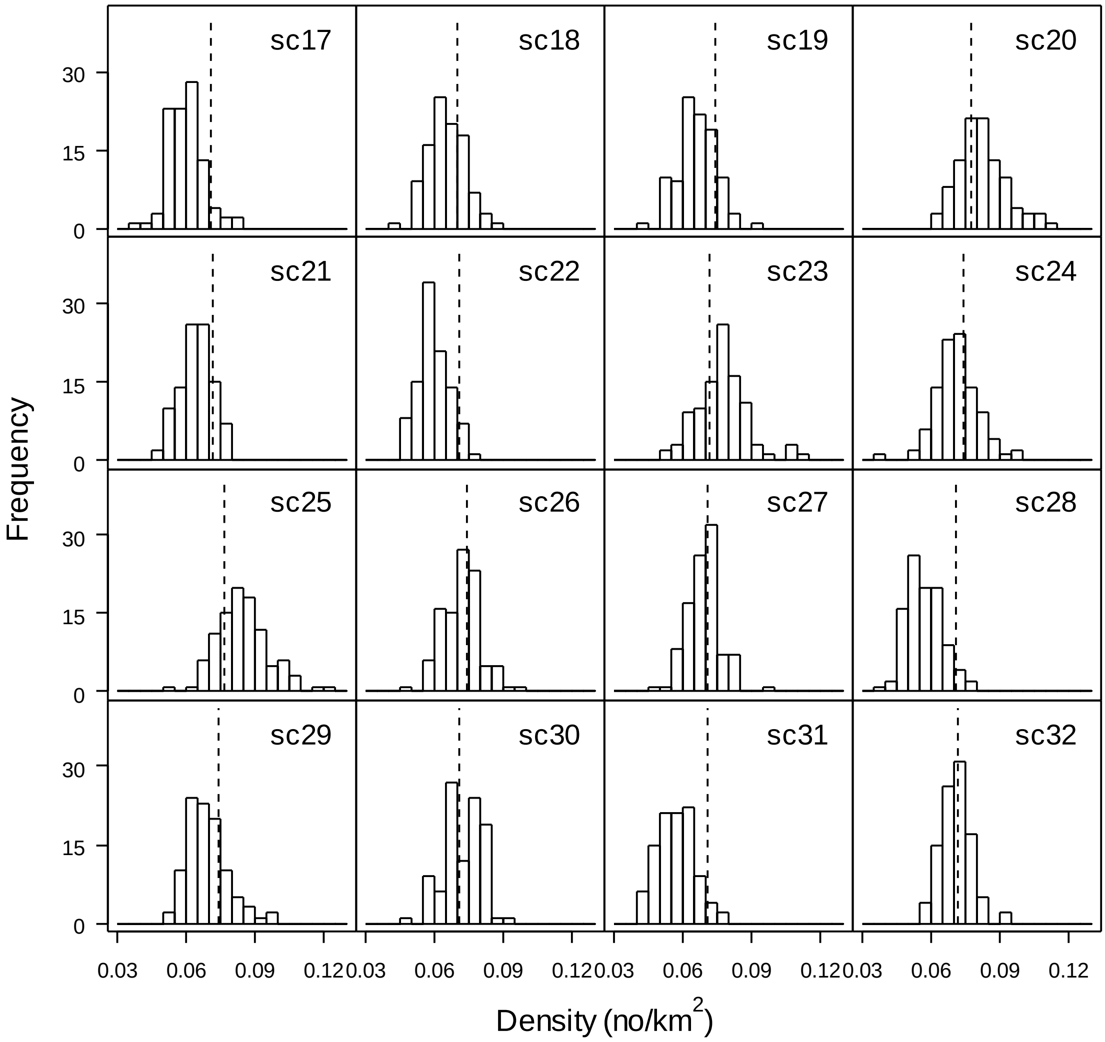

Figure 2. Histograms of estimated density (whales per km^2) for the each of the “2005” scenarios. The true mean values in the simulations are indicated by dashed vertical lines.