

- |  |   |
|--|---|
| <ul style="list-style-type: none"> <li>3.2.3.2 Preliminary investigation of experimental ways to distinguish among competing hypotheses</li> <li>3.2.3.3 Dispersal rates</li> <li>3.2.3.4 Abundance estimates (including <math>g(0)</math> issues and plans for future surveys)</li> <li>3.2.3.5 Catch data (including alternative series)</li> <li>3.2.3.6 Future whaling operations</li> <li>3.2.3.7 Other anthropogenic removals</li> </ul> | <ul style="list-style-type: none"> <li>3.2.3.8 Conditioning (including biological parameters)</li> <li>3.2.4 Work plan</li> </ul> |
| <ul style="list-style-type: none"> <li>4. Consideration of surveys under the guidelines and requirements           <ul style="list-style-type: none"> <li>4.1 Norwegian surveys</li> <li>4.2 Other</li> </ul> </li> <li>5. Other business</li> <li>6. Adoption of report</li> </ul>  |   |

## Appendix 2

### EVALUATING CRITERIA FOR DEFINING CONSERVATION PERFORMANCE FOR IMPLEMENTATION SIMULATION TRIALS (ISTs)

Andre E. Punt and Cherry Allison

IWC (2006) proposed the following steps to determine the conservation performance of an RMP variant for each stock in an *IST* for which  $MSYR=1\%$ .

- (1) Construct a single stock trial, which is 'equivalent' to the *Implementation Simulation Trial*. For example, if a particular *IST* involved carrying capacity halving over the 100-year projection period, the 'equivalent single stock trial' will also involve carrying capacity halving over the next 100 years.
- (2) Conduct two sets of 100 simulations based on this single stock trial in which future catch limits are set by the Catch Limit Algorithm (*CLA*). The two sets of simulations correspond to the 0.6 and 0.72 tunings of the *CLA*. Rather than basing these calculations on a single initial depletion, the simulations for each stock to be conducted for the set of initial depletions for the stock concerned in the *ISTs*.
- (3) The cumulative distributions for the final depletion and for the depletion ratio (the minimum over each of the 100 years of the ratio of the population size to that when there are only incidental catches) to be constructed for each tuning of the *CLA*.

The lower 5%ile of these distributions to form the basis for determining whether the performance of the RMP for the *IST* is 'acceptable', 'borderline' or 'unacceptable'.

Fig. 1 outlines these steps for a hypothetical case. Fig. 1(a) shows the cumulative distribution of the initial depletion in the *IST*. Figs 1(b) and (c) show cumulative distributions for the final depletion and the depletion ratio. These figures are constructed by running single-stock trials where the initial depletions are set to those values on which Fig. 1(a) was based. The horizontal lines in Figs 1(b) and (c) are used to compute the thresholds that define 'acceptable', 'borderline' or 'unacceptable' performance. Note that the performance of an RMP variant for a specific *IST* is 'acceptable' if either the final depletion or the depletion ratio satisfies the criteria for 'acceptable', and similarly for 'borderline' (Table 1).

At its 2005 meeting, the Committee requested that three analyses be conducted to evaluate the approach outlined above. The results of this work done intersessionally are reported in SC/58/RMP2 and described below.

#### CUMULATIVE DISTRIBUTIONS FOR SINGLE-STOCK TRIALS

Fig. 2 shows cumulative distributions for the initial depletion, the final depletion, and the depletion ratio for six initial depletion distributions (two cases for mean initial depletions of 0.3, 0.5 and 0.7) to illustrate the sensitivity of the thresholds. The difference between the two cases is the inter-simulation variation in the initial depletion (standard deviations of 0 and 0.05; 0.05 was chosen because it is large enough to differ from 0 and small enough that the impact of the mean initial depletion is still evident). Table 2 lists the values for the thresholds. Note that the initial depletion for the trial in which there is only a single initial depletion is set equal to the mean of the initial depletions for the trial in which initial depletion varied among simulations.

The values for the thresholds that define 'acceptable' and 'borderline' performance for the two performance statistics are, as expected, most sensitive to the initial depletion level. However, the values of these thresholds are fairly insensitive to variation in the initial depletion, except when the initial depletion is low ( $\sim 0.3$ ). This result is not unexpected given that the standard deviation of the initial depletion distribution was pre-specified when constructing the trials for which results are reported in Fig. 2 and Table 2.

#### APPLICATION TO *ISTs* DEVELOPED FOR NORTH ATLANTIC MINKE WHALES

Table 3 lists the initial depletions for all the stocks in all the North Atlantic minke whale trials for which  $MSYR=1\%$ , the resulting thresholds for the final depletion and depletion ratio statistics, the lower 5<sup>th</sup> percentiles for the final depletion and depletion ratio based on the actual trials, and the net result of using the criteria developed by IWC (2006) to evaluate conservation performance. The catch limits for the North Atlantic minke whale trials are set by catch-cascading over the central and eastern *Medium Areas* (IWC, 1994). These results of the North Atlantic minke trials are not identical to those in IWC (1993) because, for consistency with the single-stock trials, all of the calculations are based on the *CATCHLIMIT* program. IWC (2006) agreed that this program would be used in simulation trials owing to improvements in computing speed. However, for completeness, the results based on the old *CLA* program are included in parentheses in Table 3. As expected,

differences between the results based on the old *CLA* program and on *CATCHLIMIT* are small.

The results in Table 3 indicate that the conservation performance of the RMP variant selected for the North Atlantic minke whales would be classified as 'acceptable' for the *ISTs* considered in Table 3 (the conservation performance of a trial is taken to be that for the stock for which conservation performance is evaluated to be poorest). There were two cases in which performance on one of the two performance statistics was below the threshold that defines 'acceptable' performance, but overall performance was 'acceptable'.

#### APPLICATION TO *ISTs* DEVELOPED FOR NORTH PACIFIC MINKE WHALES

IWC (2006) suggested calculating the thresholds for the four baseline trials for the North Pacific minke whales (IWC, 2003). Table 4 lists the initial depletions for the 'J', 'O' and 'W' stocks, the resulting thresholds for the final depletion and depletion ratio statistics, the lower 5th percentiles for the final depletion and depletion ratio based on the actual trials, and the net result of using the criteria developed by IWC (2006) to evaluate conservation performance. Note that 'zero catch' trajectories on which the depletion ratio is based assume no RMP catches, although there are incidental catches, as specified for the trials concerned. The catch limits for the North Pacific minke whales are set by treating sub-areas 1, 2, 3, 4, 5, 6, 9N, 10 and 13 as *Residual Areas*, and sub-areas 7+8+11+12 and sub-area 9 as combination areas with catch limits cascaded to sub-areas within each combination area. The results for the *ISTs* differ from those reported in IWC (2003) because they are based on the *CATCHLIMIT* program. However, for completeness, the results based on the old *CLA* program are

included in a parenthesis in Table 4. As expected, there are few differences between the results based on old *CLA* program and on *CATCHLIMIT*.

The results in Table 4 indicate that the conservation performance of the RMP variant selected for North Pacific minke whales would be classified as 'acceptable' for the baseline trials based on stock structure hypotheses A and B and 'borderline' for the baseline trials based on stock structure hypotheses C and D. The conservation performance for the 'J' stock is 'acceptable' even though the final depletion is low. This is because there are few RMP catches from this stock so that depletion ratio is close to 1 (i.e. the population trajectory with RMP catches is similar to that without RMP catches). The primary reason for the low final depletion values for the 'J' stock in Table 4 is that the thresholds are calculated from single stock trials that do not include bycatch, whereas the multi-stock trials do include bycatch. It is not clear how to include bycatch in the single stock trials because bycatch is taken from multiple stocks but the use of the depletion ratio as a performance measure obviates the need for further consideration of this.

#### CONCLUSION

The sub-committee noted that there is still a fair amount of inter-simulation variability in the values for the thresholds. It may be appropriate therefore to conduct more than one simulation for each initial depletion when initial depletion varies among simulations in a trial. The sub-committee also noted, while conducting these calculations, the authors had discovered that *MSYR* had been specified in several different ways in previous *ISTs*. Standardising this to specifications that are already implemented in the single-stock control program would reduce the workload needed to apply the criteria for evaluating conservation performance.

Table 1  
Categories of conservation performance.

Acceptable	Borderline	Unacceptable
Either: lower $v$ %-ile of final depletion $> \alpha K$ ; and/or the lower $x$ %-ile of the minimum over each of the 100 years of the ratio of the population size to that in the same scenario but there are only incidental catches, remains $> \gamma$ .	Either: lower $w$ %-ile of final depletion $> \beta K$ ; and/or the lower $y$ %-ile of the minimum over each of the 100 years of the ratio of the population size to that in the same scenario but there are only incidental catches, remains $> \delta$ .	Anything that cannot be classified as 'acceptable' or 'borderline'

Table 2

Values for the thresholds that define 'borderline' and 'acceptable' performance for three average initial depletion levels and two values for the standard deviation of the initial depletion level. 'Borderline' thresholds are calculated using the 0.6 tuning of the *CLA*. 'Acceptable' thresholds are calculated using the 0.72 tuning of the *CLA*. *MSYR*=1% for all trials.

Average initial depletion	SD initial depletion	'Borderline' thresholds		'Acceptable' thresholds	
		Lower 5%ile Final depletion	Depletion ratio	Lower 5%ile Final depletion	Depletion ratio
0.3019	0	0.425	0.607	0.547	0.795
0.3019	0.05	0.388	0.578	0.436	0.763
0.5018	0	0.437	0.500	0.606	0.693
0.5018	0.05	0.435	0.500	0.601	0.682
0.7051	0	0.448	0.468	0.621	0.649
0.7051	0.05	0.439	0.466	0.617	0.641

Table 3

Summary of the application of the performance criteria for the North Atlantic minke trials. The two numbers for each threshold are respectively those based on the 0.6 and 0.72 tunings of the *CLA*. The values in parenthesis in the 'Lower 5%ile' columns are the results based on the using the *CLA* program on which the actual trials were based. Values indicated in bold-underline are not 'acceptable' for one of the two performance statistics.

Trials/stock	Initial depletion	Final depletion		Depletion ratio		Conservation performance
		Thresholds	Lower 5%ile	Thresholds	Lower 5%ile	
<b>Trial NO-1</b>						
CIC	0.7628	0.396/0.562	0.642 (0.642)	0.418/0.593	0.685 (0.685)	Acceptable
CM	0.8195	0.403/0.560	0.622 (0.623)	0.417/0.582	0.664 (0.664)	Acceptable
EN	0.5269	0.383/0.543	0.574 (0.592)	0.459/0.653	0.741 (0.763)	Acceptable
EC	0.5009	0.385/0.541	0.543 (0.569)	0.469/0.665	0.666 (0.681)	Acceptable
ES	0.4690	0.383/0.536	0.585 (0.598)	0.483/0.679	0.754 (0.771)	Acceptable
EB	0.4759	0.385/0.536	0.583 (0.597)	0.480/0.676	0.751 (0.771)	Acceptable
<b>Trial NO-3</b>						
EN	0.4442	0.385/0.530	0.542 (0.560)	0.500/0.691	0.793 (0.817)	Acceptable
EC	0.4168	0.386/0.523	0.523 (0.545)	0.510/0.704	0.736 (0.759)	Acceptable
ES	0.3809	0.380/0.515	0.546 (0.561)	0.534/0.727	0.799 (0.820)	Acceptable
EB	0.3877	0.383/0.512	0.545 (0.563)	0.529/0.721	0.797 (0.822)	Acceptable
<b>Trial NO-4</b>						
CIC	0.7858	0.400/0.564	0.627 (0.627)	0.419/0.591	0.672 (0.672)	Acceptable
CM	0.8084	0.403/0.564	0.616 (0.620)	0.417/0.586	0.658 (0.662)	Acceptable
<b>Trial NO-5</b>						
EN	0.6102	0.384/0.550	0.596 (0.605)	0.434/0.623	0.704 (0.715)	Acceptable
EC	0.5865	0.381/0.549	0.552 (0.564)	0.437/0.631	<b>0.602</b> (0.616)	Acceptable
ES	0.5595	0.383/0.545	0.610 (0.619)	0.448/0.638	0.720 (0.731)	Acceptable
EB	0.5659	0.383/0.546	0.606 (0.615)	0.446/0.636	0.716 (0.727)	Acceptable
<b>Trial NO-6</b>						
CIC	0.7271	0.393/0.561	0.611 (0.616)	0.420/0.600	0.654 (0.658)	Acceptable
CM	0.8090	0.404/0.564	<b>0.561</b> (0.567)	0.417/0.587	0.591 (0.585)	Acceptable

Table 4

Summary of the application of the performance criteria for the four baseline trials for North Pacific minke trials. The two numbers for each threshold are respectively those based on the 0.6 ('borderline') and 0.72 ('acceptable') tunings of the *CLA*. Thresholds are calculated using single stock trials, which do not include bycatch. The values in parenthesis in the 'Lower 5%ile' columns are the results based on using the *CLA* program on which the actual trials were based. Values indicated in bold-underline are not 'acceptable' for one of the two performance statistics.

Case/stock	Initial depletion (median)	Final depletion		Depletion ratio		Conservation performance
		Thresholds	Lower 5%ile	Thresholds	Lower 5%ile	
<b>Baseline A</b>						
J	0.300	0.354/0.462	<b>0.111</b> (0.111)	0.587/0.764	0.964 (0.964)	Acceptable
O	0.704	0.391/0.555	0.719 (0.723)	0.407/0.593	0.751 (0.744)	Acceptable
W	0.993	0.423/0.558	0.919 (0.920)	0.409/0.554	0.908 (0.907)	Acceptable
<b>Baseline B</b>						
J	0.300	0.354/0.462	<b>0.111</b> (0.111)	0.587/0.764	0.964 (0.964)	Acceptable
O	0.728	0.394/0.557	0.728 (0.729)	0.409/0.589	0.752 (0.757)	Acceptable
<b>Baseline C</b>						
J	0.300	0.354/0.462	<b>0.108</b> (0.108)	0.587/0.764	0.968 (0.967)	Acceptable
Ow	0.253	0.347/0.457	<b>0.430</b> (0.427)	0.554/0.740	<b>0.693</b> (0.696)	Borderline
Oe	0.662	0.387/0.555	0.571 (0.565)	0.406/0.595	0.723 (0.731)	Acceptable
W	0.990	0.408/0.555	0.819 (0.824)	0.398/0.552	0.817 (0.822)	Acceptable
<b>Baseline D</b>						
J	0.300	0.354/0.462	<b>0.109</b> (0.109)	0.587/0.764	0.969 (0.969)	Acceptable
O	0.290	0.356/0.482	<b>0.381</b> (0.385)	0.510/0.699	<b>0.692</b> (0.688)	Borderline
W	0.987	0.408/0.554	0.804 (0.806)	0.398/0.554	0.797 (0.800)	Acceptable

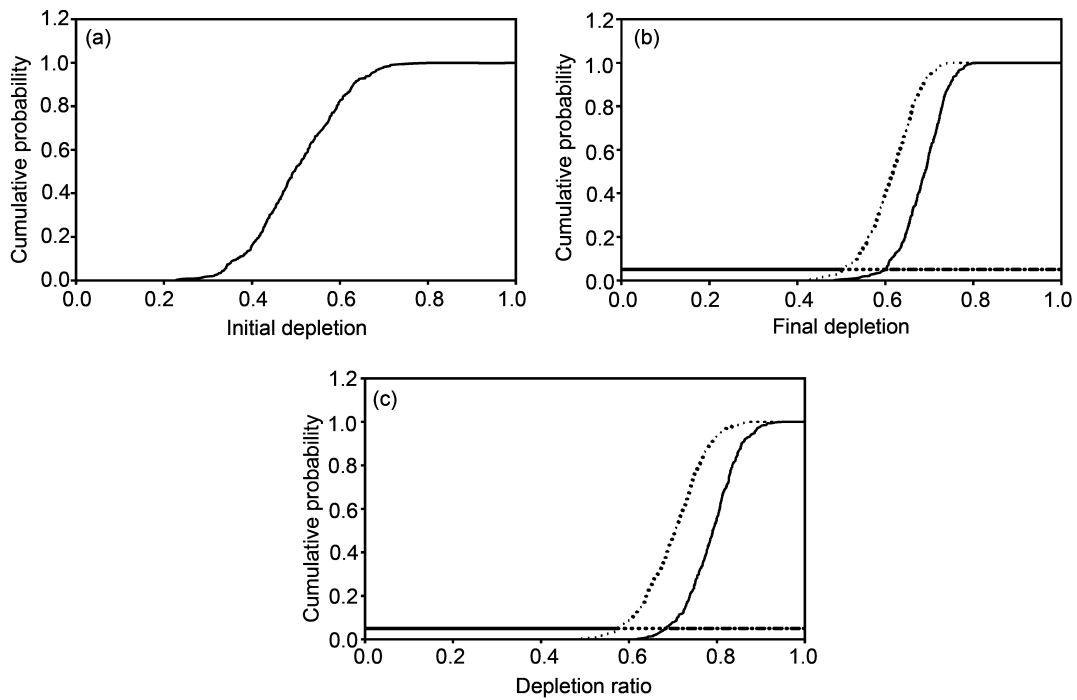


Fig. 1. Application of the performance criteria evaluated in this paper to a hypothetical *IST*. Panel (a) plots the distribution for the initial depletions in the trial. The solid and dotted lines in panels (b) and (c) denote the cumulative distributions for the 0.72 and 0.60 tunings of the *CLA*. The solid, dotted and dashed horizontal lines denote the ranges for the two performance statistics for which performance would be considered to be ‘unacceptable’, ‘borderline’ and ‘acceptable’ respectively.

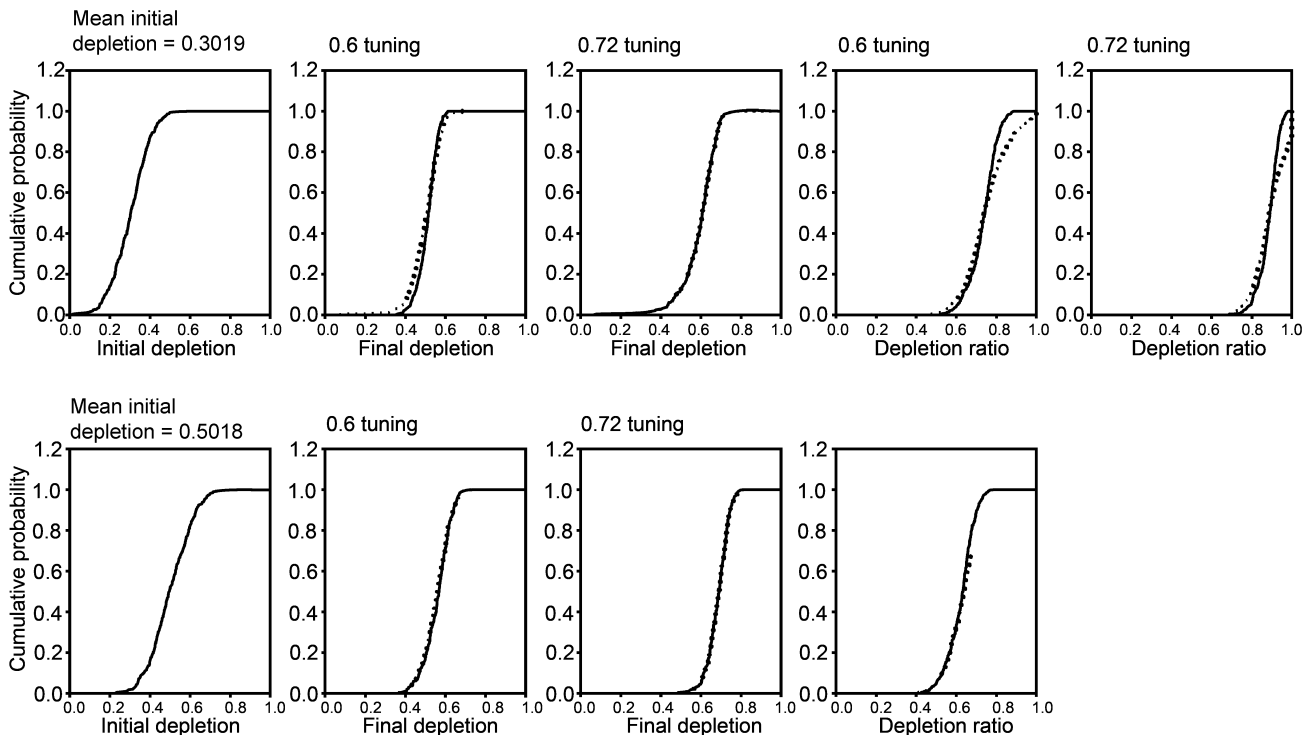


Fig. 2. Initial depletion distribution, distributions for the final depletion and the depletion ratio for the 0.6 tuning of *CLA*, and distributions for the final depletion and the depletion ratio for the 0.72 tuning of *CLA*. Results are shown for three mean initial depletions. The solid and dotted lines in columns 2-4 denote respectively results when there is no inter-simulation variation in the initial depletion and when the inter-simulation standard deviation of the initial depletion is 0.05.

REFERENCES

International Whaling Commission. 1993. Report of the Scientific Committee, Annex I. Report of the Working Group on Implementation Trials. *Rep. int. Whal. Commn* 43:153-96.  
 International Whaling Commission. 1994. Report of the Scientific Committee, Annex D. Report of the Sub-Committee on Management Procedures. *Rep. int. Whal. Commn* 44:74-92.

International Whaling Commission. 2003. Report of the Scientific Committee, Annex D. Report of the Sub-Committee on the Revised Management Procedure. *J. Cetacean Res. Manage. (Suppl.)* 5:107-53.  
 International Whaling Commission. 2006. Report of the Scientific Committee, Annex D. Report of the Sub-Committee on the Revised Management Procedure (RMP). *J. Cetacean Res. Manage. (Suppl.)* 8:78-90.