Report of the Scientific Committee

Bled, Slovenia, 7-19 June 2016

Annex T: Matters related to discussions of NEWREP-A

**This report is presented as it was at SC/66b.
There may be further editorial changes (e.g. updated references, tables, figures) made before publication.**

International Whaling Commission

Bled, Slovenia, 2016

Annex T

Matters related to discussions of NEWREP-A

ANNEX T1: REPORT OF THE *AD HOC* GROUP ON SCAA EVALUATION FOR SOUTHERN HEMISPHERE MINKE WHALES

PUNT (CHAIR), BUTTERWORTH, COOKE, DE LA MARE, KITAKADO, MCKINLAY, WALLØE

SC/66b/IA08 concluded based on the results of simulations that there is little ability to estimate MSYR for Southern Hemisphere minke whales even if a Statistical Catch-at-age (SCAA) model is correctly specified. In contrast, applications of the SCAA developed by Punt *et al.* (2014) suggest that there is an appreciable difference in the value of the objective function among different values of MSYR (Table 1; Total).

The results in Table 1 indicate that the bulk of the difference in objective function value between the model scenarios based on MSYR=1% (20.66 out of 28.90) and MSYR=2.5% (6.86 out of 9.70) and that based on MSYR=4% is due to penalties on 'random effects'[1] parameters in the stock-recruitment relationship (the recruitment and carrying capacity penalties). Some of the change in the objective function value for the conditional age-at-length data among MSYR values may also be due to the penalties on the recruitment and carrying capacity as conditional age-at-length data provide information on year-class strength, which is related to how recruitment is modelled. However, there was insufficient time to conduct analyses to investigate this.

The key feature of the data for Southern Hemisphere minke whales is a change in age composition over the early years of the fishery. In the SCAA developed by Punt *et al.* (2014) this is attributed to age-specific natural mortality, and changes over time in selectivity, recruitment deviations (see Adjunct 1) and carrying capacity. However, it could also be due (in whole or in part) to changes over time in natural mortality. Furthermore, the impact of changes in carrying capacity have a lesser impact when MSYR is lower than when MSYR is higher given the recruitment function in the Punt et al. (2014) SCAA and that of SC/66b/IA08.

The Small Group **recommends** that the following analyses be undertaken to further explore this issue:

- decrease the effect of the penalties on the recruitment and carrying capacity deviations to understand whether these penalties (which are not imposed in SC/66b/IA08) are the main reason for the apparent discrimination ability of the Punt *et al.* SCAA method; and
- extend the SCAA to include density-dependent natural mortality. The Siler model approach in SC/66b/IA08 is one way to account for time-varying natural mortality.

REFERENCE

Punt, A.E., Bando, T., Hakamada, T. and Kitakado, T. 2014. Assessment of Antarctic minke whales using Statistical Catch-at-age Analysis. Journal of Cetacean Research and Management. 14: 93–116.

Table 1

Difference in the objective function as a function of MSYR

Objective function component	MSYR		
	1%	2.5%	4%
Total	28.90	9.70	0.00
Absolute abundance	-0.01	0.02	0.00
Relative abundance	0.68	0.26	0.00
Total catch	0.00	0.00	0.00
Length frequency	0.48	0.39	0.00
Conditional age-at-length	7.70	2.40	0.00
Recruitment penalty	15.65	4.82	0.00
Selectivity penalty	0.00	0.00	0.00
Carrying capacity penalty	5.01	2.04	0.00
Stock proportion penalty	-0.18	-0.09	0.00
Growth penalty	-0.48	-0.10	0.00

[1] These are not true random effects because the random effects variance is pre-specified rather than being estimated.

ANNEX T2. REPORT OF THE AD HOC GROUP ON REVIEWING RECOMMENDATIONS RELATED TO MODELLING

PUNT (CHAIR), BUTTERWORTH, COOKE, DE LA MARE, KITAKADO, MCKINLAY, WALLØE

SC/66b/SP10 provides additional analyses and information related to analytical aspects for six of the recommendations of the Expert Panel that reviewed NEWREP-A. The Small Group considered the progress made in addressing each of these recommendations. Text developed by the small group for the overall summary of progress is included in the Table X under Plenary Item 18.1 and so is not repeated here while specific detailed comments and suggestions related to recommendations 1 and 26 given below:

Recommendation 1.

(1) It is hard to compare the performances of the CLA and MCLA as they are tuned differently. The factor that multipliers the output of the CLA in the MCLA should be chosen so that that the performance (catch or depletion) is that same for the CLA and MCLA for a selected 'reference' trial.

(2) The proponents may wish to report additional performance statistics.

(3) Scenarios S1 and S2 are somewhat arbitrary. The scenarios should next base the extent of change in recruitment (condition it) on the results of the SCAA. This could be achieved, for example, by assuming that the past changes in carrying capacity and/or growth could occur in the future.

Recommendation 26.

(1) SC/66b/SP10 restricts the data used to fit the models to ages 4-13 and 1980-87 and 1992-99. All of the data should be used to estimate the amount of extra- age, -cohort and -year variation rather than restricting the analysis to a subset of years and ages. Doing this also avoids the need to simulate the process of excluding some cohorts and ages when analysing future (simulated) age data.

(2) The estimates in SC/66b/SP10 of the variance of cohort random effects and extra-binomial variation (i.e. overdispersion) are zero, which makes these asymptotic estimates potentially questionable. Use a method (such as likelihood profile or the R package *blme*) to better quantify the uncertainty of these variances and develop probability distributions for them. The simulations to evaluate power should then sample from these distributions.

(3) The current analyses do not attempt to specifically quantify the effects of year-to-year sampling variation, which reflects the impact of, for example, the locations of sampling (for examples, in some years in regions where mature animals predominate) although overdispersion arising from this sort of heterogeneity was considered to some extent in SC/66b/SP10 in beta-binomial model. Though challenging, simultaneous estimation of random effects of year and cohort can be explored using the type of model used to estimate cohort random variation in SC/66b/SP10.

ANNEX T3. ON THE ROLE OF AGE DATA IN THE PERFORMANCE OF THE MODIFIED CLA INTRODUCED IN SC/66B/SP10

KITAKADO AND BUTTERWORTH

This Annex responds to requests in Annex T2 and the some Committee members for further explanation of the results summarised in Table 3 of SC/66b/SP10, and in particular how they might indicate the advantages, if any, of including age data in a modified version of the RMP's *CLA*. This is named MCLA (modified CLA) and defined by the formula given on page 8 of SP10.

$$MCLA_t = 0.9\min(\max(I_t, 0.8), 1.2)^2 \cdot CLA_t$$

$$\text{where } I_t = \frac{\sum_{i=t-5}^{t-1}\sum_{a=1}^{5} C_{ia} \Big/ \sum_{i=t-5}^{t-1}\sum_{a=1}^{A} C_{ia}}{\sum_{i=-10}^{-1}\sum_{a=1}^{5} C_{ia} \Big/ \sum_{i=-10}^{-1}\sum_{a=1}^{A} C_{ia}}.$$

Inferences about advantages are, however, confounded by the basic trade-off between catch and depletion in any trial of the nature implemented in SC/66b/SP10. The higher the catch, the lower the population depletion – a trade-off that is controlled by a control (tuning) parameter in the MCLA – in this instance the multiplier 0.9 above. For the same value of this tuning parameter, the trade-off and how it relates to the corresponding trade-off for the *CLA* itself will vary trial by trial, rendering a comparison between the *CLA* and the MCLA difficult. For this reason, Annex T2 had suggested results be shown for a value of the control parameter 'tuned' so that one performance statistic for the MCLA had the same value as for the *CLA* for a particular trial (this is similar to the way the *CLA* is tuned to attain a specific depletion under the D1 trial).

The results are shown in Table 1 below. Table 1a duplicates the results for trials Tr1 to Tr6 which are reported in Table 3 of SC/66b/SP10. In Table 1b, the corresponding results are reported for a retuning of the MCLA which selects the value of the tuning parameter to ensure identical results under the *CLA* and MCLA for the median average annual catch of 1224 for trial Tr1. This value (for the multiplier) turns out to be 0.8.

Inspection of the results in Table 1 makes clear that as one moves from results in Table 1a with a tuning parameter of 0.9 to those in Table 1b corresponding to a value of 0.8, there is a general trend of reduced catches and higher lowest depletions for the MCLA. In practice, since it is impossible to find a tuning parameter value that achieves the same trade-off for every trial, a selection is made that is seen to provide the best trade-off across all the trials considered (note that a comprehensive selection process would involve a wider variety of trials than considered here). In principle, a different tuning parameter could be adopted for MCLAs applied to different areas so as to get a better performance overall.

Nevertheless, it is not straightforward to separate the effects of adding age data to the *CLA* and the effectively different tunings of the trade-off between catch and depletion for the other trials under the MCLA and *CLA*. To attempt such a separation, Table 2 shows results for a differential effect that aims to eliminate the effect of the effectively different tunings amongst trials. This is achieved by comparing median average annual catch results for trials with especially good recruitment with the results for those with especially poor recruitment. Under the *CLA*, these catches respectively increase or decrease compared to the corresponding trial with no such special feature in recruitment, as one would hope in response to the better or worse population abundance that follows such an up or down pulse in recruitment. The MCLA trial results exhibit the same behaviour. What is important, however, is that the difference in catches for good *vs* poor recruitment increases when changing from the *CLA* to the MCLA. This applies for both stocks, both MSYR values, and both MCLA tunings considered. This effect ranges from some 5% to over 100%, and reflects the benefits of including age data in the *CLA* through an approach that effectively removes the confounding caused by different effective tunings.

Table 1

Results comparing the performances under trials Tr1 to Tr6 set out in SC/66b/SP10 for two "tunings: first in (a) for the 0.9 tuning reported in Table 3 of that paper, and secondly in (b) for an alternative tuning of 0.8 which leads to the same median average annual catch under the CLA and MCLA for trial Tr1.

(a) Tuning paramter = 0.9

I-stock			Median average catch				Lower 5%-ile of lowest depletion			
Trial	MSYR1+	Rercruitment Scenario	MCLA	CLA	Difference	Performance improvement	MCLA	CLA	Difference	Performance improvement
Tr1	2.5%	S0 (flat)	1323	1224	99	Y	0.700	0.723	-0.023	-
Tr2	2.5%	S1(poor)	861	817	44	Y	0.472	0.448	0.024	Y
Tr3	2.5%	S2 (good)	1527	1415	112	Y	0.721	0.742	-0.021	-
Tr4	4%	S0 (flat)	1599	1438	161	Y	0.760	0.787	-0.027	-
Tr5	4%	S1(poor)	1223	1174	49	Y	0.537	0.530	0.007	Y
Tr6	4%	S2 (good)	1751	1567	184	Y	0.770	0.797	-0.027	-

P-stock			Median average catch				Lower 5%-ile of lowest depletion			
Trial	MSYR1+	Rercruitment Scenario	MCLA	CLA	Difference	Performance improvement	MCLA	CLA	Difference	Performance improvement
Tr1	2.5%	S0 (flat)	1118	1003	115	Y	0.693	0.728	-0.035	-
Tr2	2.5%	S1(poor)	826	765	61	Y	0.457	0.466	-0.009	N
Tr3	2.5%	S2 (good)	1328	1169	159	Y	0.712	0.754	-0.042	-
Tr4	4%	S0 (flat)	1363	1206	157	Y	0.736	0.762	-0.026	-
Tr5	4%	S1(poor)	1223	1174	49	Y	0.508	0.512	-0.004	N
Tr6	4%	S2 (good)	1508	1338	170	Y	0.763	0.797	-0.034	-

(b) Tuning paramter = 0.8

I-stock			Median average catch				Lower 5%-ile of lowest depletion			
Trial	MSYR1+	Rercruitment Scenario	MCLA	CLA	Difference	Performance improvement	MCLA	CLA	Difference	Performance improvement
Tr1	2.5%	S0 (flat)	1224	1224	0	-	0.717	0.723	-0.006	-
Tr2	2.5%	S1(poor)	780	817	-37	N	0.492	0.448	0.044	Y
Tr3	2.5%	S2 (good)	1401	1415	-14	N	0.762	0.742	0.020	Y
Tr4	4%	S0 (flat)	1434	1438	-4	-	0.781	0.787	-0.006	-
Tr5	4%	S1(poor)	1116	1174	-58	N	0.547	0.530	0.017	Y
Tr6	4%	S2 (good)	1570	1567	3	-	0.797	0.797	0.000	-

P-stock			Median average catch				Lower 5%-ile of lowest depletion			
Trial	MSYR1+	Rercruitment Scenario	MCLA	CLA	Difference	Performance improvement	MCLA	CLA	Difference	Performance improvement
Tr1	2.5%	S0 (flat)	1040	1003	37	Y	0.705	0.728	-0.023	-
Tr2	2.5%	S1(poor)	785	765	20	Y	0.466	0.466	0.000	-
Tr3	2.5%	S2 (good)	1228	1169	59	Y	0.746	0.754	-0.008	-
Tr4	4%	S0 (flat)	1265	1206	59	Y	0.759	0.762	-0.003	-
Tr5	4%	S1(poor)	1031	1174	-143	N	0.526	0.512	0.014	Y
Tr6	4%	S2 (good)	1409	1338	71	Y	0.781	0.797	-0.016	-

Table 2.

Differences in median average catch between good (S2) and poor (S1) recruitment scenarios for the MCLA compared to CLA. The first two figures shown in each equation are the differences between the catch results for the S2 and S1 scenarios for first the MCLA, and then followed for the CLA. Thus for example in Table 1(a) the Tr3 (S2) trial shows a catch result of 1527 while the Tr2 (S1) trial shows a corresponding result of 861 for the MCLA with a 0.9 tuning. The difference between these two figures is 666 as shown in the corresponding initial entry below.

Stock	MSYR1+	Tuning paramter	
		0.9	0.8
I-stock	2.5%	666 - 598 = 68	621 - 598 = 23
	4%	528 - 393 = 135	454 - 393 = 61
P-stock	2.5%	502 - 404 = 98	443 - 404 = 39
	4%	285 - 164 = 121	378 - 164 = 214

Adjunct 1: Further clarifications regarding the role of age data in and the performance of the modified CLA introduced in SC/66b/SP10

It may be useful to try to clarify some matter that arose in discussions about SC WP07.

The primary intent of SC/66b/SP10 was to demonstrate the utility of the collection of age data for inclusion in a modification of the CLA that improves its performance. Table 3 of that document, which is near duplicated by Table 1(a) of Annex X3, provided a summary of results to support that demonstration.

Two separate issues need to be distinguished. The first is the role that the age data may play in the argued improvement of the *CLA* when it is adjusted to include age data (the MCLA). Table 2 of Annex T3 was provided to address that point. In that table, an attempt was made to remove the confounding effect of different catch vs depletion trade-offs when making comparisons.

The other issue concerns whether the performance of the MCLA overall reflects an improvement over that of the *CLA*. Table 3 of SC/66b/SP10 was intended as an illustration in support of the general contention that including age data in some MCLA can result in improved performance overall. It was noted there that in nearly all cases, the MCLA led to either or both of catch being increased and the lower 5%ile levels of lowest depletion being improved (where necessary) compared to the *CLA*. In the two instances (indicated by N) where the lower 5%iles of lowest depletion were low under the *CLA*, and further reduced under the MCLA, this reduction was not large.

As explained above, the choice for the tuning parameter associated with MCLA or any form of modified *CLA* will lead to different catch-depletion trade-offs for different trials. These could mean that MCLA catches in all these trials are lower (indicated by N) and all lower 5%iles for lowest depletion higher (Y or -) than for the *CLA*, and vice versa. At intermediate levels of the tuning parameter, some mixture of Y's and N's will result (as in the two tuning examples in Table 1 below). Like in this case, the MCLA with any specific choice of tuning parameter may not uniformly dominate the *CLA* for all trials. This implies that it is highly unlikely that what is called a 'uniformly dominant' solution will emerge for a tuning choice for which there are no N's. As was the case when the *CLA* itself was selected, a final choice would seek what is considered to be the best trade-off across the various trials. Therefore, in this case, for example, one might choose the 0.9 tuning of the MCLA for the area corresponding to the I stock, and the 0.8 tuning for that associated with the P stock, resulting in only one case with an N (for a reduced catch) under the MCLA, and no instances of worse performance for the lower 5%ile of lowest depletion.

However, this would be going beyond the illustrative intent of SC/66b/SP10. As stated there, a final choice of the tuning parameter of the MCLA, or of any form of modified *CLA*, would need to involve a much wider set of trials than considered in the illustrative example whose results are shown in Table 3 of that paper. Probably also the MCLA would be refined in that process to improve its performance further. However, for this purpose, in line with the Committee's customary practice, a pre-requisite for this further work is for the Committee to provide a pre-specified set of agreed trials (the proponents – the Government of Japan - if contributing to such further work, should not be expected to invest considerable time in developing and running further trials, only to be informed later by the Committee that they would have wanted different trials run). Accordingly, the Government of Japan states in SC/66b/SP10 that it looks forward to the Committee agreeing on the specifications of an extension to the trials undertaken in that paper (or at least, more immediately, on a process to develop those specifications in the Committee), so that work on this matter can continue efficiently to advise the Committee.

ANNEX T4. TERMS OF REFERENCE FOR AN ADVISORY GROUP ON RECOMMENDATION 1 OF THE EXPERT PANEL ON NEWREP-A

BANNISTER (CHAIR), BUTTERWORTH, COOKE, DE LA MARE, DONOVAN, FORTUNA, MCKINLAY, KITAKADO, MORISHITA, PUNT, WALLØE

Recommendation 1 by the Expert Panel on NEWREP-A is:

> 'Evaluate the level of improvement that might be expected either in the SCAA or in RMP performance by improved precision in biological parameters using simulation studies including updated *Implementation Simulation Trials*'.

SC/66b/SP10 provided such RMP/IST-like simulations to evaluate whether or not a modified CLA[2] that includes age data in the control rule will (a) result in improved performance and (b) if so by how much. The Small Group (Annex X2) had agreed that this approach is conceptually the appropriate way to respond to the Panel's recommendation, but recognised that further work is needed to specify the appropriate trial structure.

The Terms of Reference and Workplan for an Advisory Group to move forward the process of assisting the proponents to provide advice, especially with respect to mathematical specifications concerning Recommendation 1[3] are to:

(1) provide complete mathematical specifications for the trials conducted in SC/66b/SP10 and distribute them for comment by the Advisory Group. The specifications could be in the form of the existing model specifications for the SCAA as an appendix, with additional text and equations highlighting the additions to generate future data, impose future changes in recruitment and any other parameters used in setting up the trials. (Japan);

(2) provide comments on the mathematical specifications for the existing analyses (Advisory Group);

(3) identify additional factors that should be considered in trials (both those specified in 4 below and possibly others, indicating also a broad order of priority) (Advisory Group);

(4) provide the mathematical specifications for trials to account for factors identified in Annexe X2 (i.e., time-varying recruitment due to changes in carrying capacity, time-varying age-at-sexual maturity, time-varying natural mortality) and during subsequent Committee discussions (sampling of age data, mixing of the I and P stocks) and distribute them for comment by the Advisory Group. Document the intended process for conditioning the trials given these additional features, including specifications of the likelihood functions / penalties and the types of data should be specified, along with assumed parameters values where data are missing or incomplete (Japan);

(5) provide comment on the mathematical specifications for the extensions resulting from TOR 3 and 4 (Advisory Group); and

(6) identify issues to be addressed during the 2017 Committee meeting (Advisory Group).

The Advisory Group will assist in the process to facilitate the Committee to review and agree trial specifications during SC67a concerning Recommendation 1. It is recognised that the trial process is iterative. There is no expectation that Japan will be able to code and run the trials prior to SC67a.

[2] Should Japan ultimately wish to propose a modification to the *CLA,* this must be within the Committee's existing framework to evaluate modification to the CLA (ref). This framework, that requires the proposers to undertake simulation evaluations of some kind, was most recently applied to the Norwegian proposal to modify the *CLA* (refs).

[3] The existence of a range of opinions about the extent to which Recommendation (1) had been addressed was recognised in Plenary.

ANNEX T5. UNCERTAINTY IN THE OVERDISPERSION PARAMETER IN BETA-BINOMIAL MODEL ASSESSED WITH A RELATIVE LOG-LIKELIHOOD PROFILE FUNCTION

KITAKADO

SC/66b/SP10 showed that the point estimate of overdispersion in beta-binomial distribution was zero and the associated standard error was also negligible. However, concern was raised if the Fisher Information matrix-based standard error may not be appropriate to use; hence rather uncertainty based on the profile likelihood or Bayesian methods should be pursued.

In Fig. 1, the profile likelihood with respect to overdispersion is reported. There is uncertainty to some extent, but the uncertainty does not seem large. Nevertheless, further simulation for power analysis on the age-at-sexual maturity (ASM) could be re-run based on this outcome, and on further analyses with full data, if requested by the Committee.

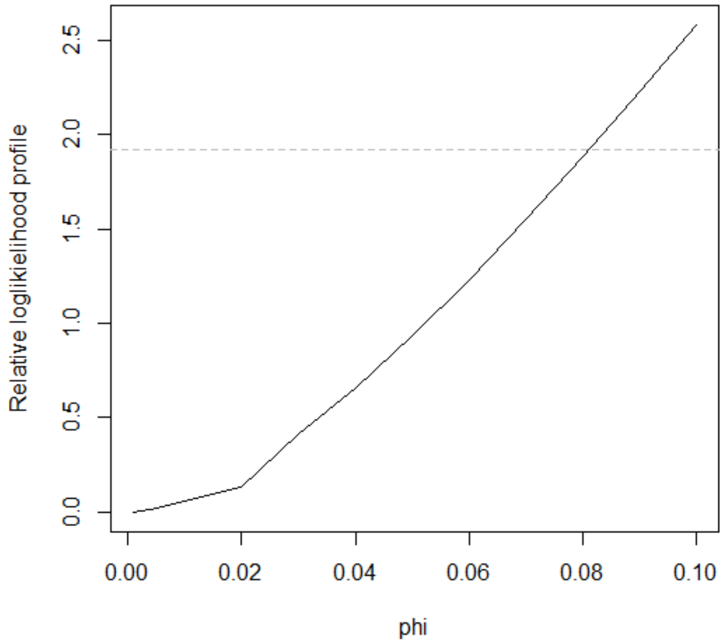

Fig. 1. Likelihood profile of the overdispersion parameter in beta-binomial model with fixed-cohort and age effects. The age range and cohorts used is same as the revised NEWREP-A proposal, Appendix 14.