Report of the Scientific Committee

Bled, Slovenia, 24 April-6 May 2018

Annex L
Report of the Working Group on Ecosystem Modelling

This report is presented as it was at SC/67b.
There may be further editorial changes (e.g. updated references, tables, figures) made before publication.

International Whaling Commission
Bled, Slovenia, 2018

Annex L

Report of the Working Group on Ecosystem Modelling

Members: Kitakado (Convenor), Al Harthi, Aoki, Baba, Bell, Brierley, Brownell, Burkhardt, Buss, Butterworth, Cañadas, Cooke, Dalla Rosa, de la Mare, de Moor, DeWoody, Di Tullio, Domit, Doniol-Valcroze, Donovan, Double, Ferguson, Ferriss, Fortuna, Galletti Vernazzani, Gunnlaugsson, Haug, Hielscher, Irvine, Kim, E., Konishi, Lang, Lundquist, Mallette, McKinlay, Miller, Morita, Moronuki, Mosquera Guerra, Murase, New, Noren, Øien, Palka, Panigada, Pastene, Pierce, Punt, Reeves, Reyes Reyes, Ritter, Rowles, Santos, Simmonds, Skaug, Slugina, Solvang, Strasser, Svoboda, Taguchi, Tamura, Taylor, Torres, Víkingsson, Wade, Walløe, Willson, Yasokawa, Yasunaga, Zerbini, Zharikov.

1. INTRODUCTORY ITEMS

1.1 Introductory remarks
Kitakado welcomed the members of the Ecosystem Modelling Working Group (hereafter Working Group).

1.2 Election of Chair
Kitakado was elected Chair.

1.3 Appointment of rapporteur
New was appointed rapporteur.

1.4 Adoption of Agenda
The adopted Agenda is included as Appendix 1.

1.5 Documents available
The documents available to the Working Group were identified as SC/67b/EM01-08, SC/67b/SP09, Cunen *et al.* (in review, SC/67b/FI02) and de la Mare *et al.* (submitted, SC/67b/FI28).

2. BODY CONDITION ANALYSES

2.1 Review progress of analyses for Antarctic minke whales
2.1.1 Review of analyses

SC/67b/EM03 reported updated analyses of minke whale body condition using data collected under the JARPA special permit whaling programme. Generalised additive models (GAMs) using penalized regression splines were used to develop models of body weight and blubber thickness (BT11), with a focus on examining the likely shape and variability in any time trends in the accumulation of weight and BT11 over the main part of the summer feeding season. In order to ameliorate space-time confounding introduced by the biennial nature of the JARPA sampling program, separate analyses were conducted for animals taken in the West and the East sampling regions. As all female animals sampled were pregnant, and different bioenergetic relationships might be expected between males and females, separate analyses were also conducted for each of the sexes in order to completely separate sex effects. The long-term trend in condition was examined using model predictions conditioned to evaluate the difference in body weight and BT11 for early- and late-season animals. This provided biennial estimates (with uncertainty) of a relative index of accumulated weight or BT11 over the main part of the summer feeding season. Results indicated very little trend or variability in the improvement in body condition attributable to the period of summer feeding sampled by JARPA, for males or females from the West or East sampling regions. Some slight evidence of a consistent decline for females in the West was reported, but the total decline over eight sampling seasons for this portion of the population was no more than the year-to-year variability found in other segments of the population sampled under JARPA. Results from models of body weight and BT11 were consistent, and models fitted well by all usual criteria. Models also revealed an important, previously unreported relationship between body length, blubber thickness and body weight, showing that blubber thickness proportionately decreases with increasing body length, provided models account for body weight. This relationship was also demonstrated using plots of the raw data, independently of the model results. In addition, the authors demonstrated that average catch lengths have increased over the period of JARPA and that, combined, the implication of these results is that failing to correctly model the length-weight-BT11 relationship will result in false signals of changing condition. The authors posit that misspecifying this important relationship is one of the reasons that other researchers have erroneously concluded body condition is declining.

SC/67b/EM02 presents an updated version of the results section from Cunen, Walløe and Hjort (2017). Compared to the report that was presented last year, the authors have made some extensions and refinements of the FIC method, and also certain moderate changes to the wide model for the JARPA data. Some of the changes are motivated by the discussions in the Scientific Committee last year. The main conclusion of the analyses is that blubber thickness, fat weight and three other similar variables have declined significantly and substantially over the 18 years of the JARPA survey. There are some differences between regions, but not between the two sexes.

Due to comments by McKinlay and de la Mare, the authors of SC/67b/EM02 have also investigated the effect of including both body length and body weight or only body weight as covariates in the analyses. The analyses gave year effect estimates a bit closer to zero, but they are all significantly different from zero.

In SC/67b/EM01Rev1 the authors provide their assessment of the work presented in SC/67b/FI02 and SC/67b/EM02 in relation to analyses of minke whale body condition conducted by Norwegian and Japanese scientists. The main issues of concern were: (i) how to correctly account for spatio-temporal confounding inherent to the biennial JARPA sampling program; (ii) how the length-weight-blubber relationship should be modelled, and how omitting weight as a covariate in models of blubber will be likely to induce spurious trends in condition; (iii) that SC/67b/FI02 presents the JARPA data as if it were well-behaved, balanced data; (iv) the wide model, upon which the final result hinges, is chosen arbitrarily by the researcher without application of any objective model selection process; and, (v) that models do not account for animal-specific indicators of time spent feeding in Antarctic waters. In relation to point (iv), results were presented that illustrated the dependency of the FIC selection on the choice of the 'wide' model, which lacks a formal criterion. The estimate of the focal parameter from the wide model acts as a strong attractor in the JARPA data because the variation in bias among candidate models is greater than changes in their variance. Some bootstrap results indicated that under hypothetical replication the probability of selecting the same best model could be low.

In SC/67b/EM08 the authors present arguments against the analyses and results presented in SC/67b/EM01 and SC/67b/EM03. Compared to analyses presented to the SC by McKinlay *et al*. (2017) and in earlier years, they have this year used a new dependent variable for blubber thickness, that is the change in blubber thickness in each year "that can be attributed to summer feeding in Antarctica". To select whales that, in their opinion, fulfil this criterion they use only whales with low diatom load (males) or small foetuses (females) early in the summer season and animals with high diatom load or large foetuses late in the season. The difference in blubber thickness between these two groups was then used as a dependent variable in their analysis. However, there are large uncertainties connected to both of these indicators. Especially for the size of foetuses late in the feeding season, is possible that females with small foetuses have experienced bad feeding conditions resulting in a delayed oestrus and/or a slow growth of the foetus. The result could well have been that McKinlay *et al*. select away from their analyses whales that have experienced bad feeding conditions during the summer. Their analysis also excludes the possibility that whales that have experienced a bad feeding season in one year start out the next year with lower than average fat stores. Thus, their new dependent variable does not account for possible accumulated changes in fat stores over the years. Some of their plots of results from the GAM analyses show indications of overfitting. Even so, their results for blubber thickness are largely consistent with the results presented for blubber thickness (BT11) in SC/67b/EM02, thus indicating a decrease in blubber thickness from the beginning to the end of the JARPA period.

2.1.2 Discussion

Working Group participants were provided with a brief history on the debate surrounding the body condition analysis for Antarctic minke whales. Initial discussion began in 2014, with the conclusion that there had been a statistically significant (5% level) decline in blubber thickness and fat weight [IWC 2015]**.** Since then there has been some collaboration and considerable development in the types of models used, as well as in-depth discussions regarding the proper handling of data and the explanatory variables to be included in the analysis. Generally, the differences in results are due to the different use of data and models, as well as some areas of statistical debate. In addition, this year a new variable of primary interest was introduced in SC/67b/EM03 by McKinlay, de lar Mare and Welsh (MDW). Cunen, Walløe and Hjort (CWH) consider that this variable was not as appropriate as those used previously. Tables 1 and 2 provide summaries of the comparisons of the model specification and inferences for the models presented by MDW and CWH.

When discussing SC/67b/EM03, it was noted that there are two ways to implement the shrinkage approach when using generalized additive models. The one taken in SC/67b/EM03 would imply the shrinking of the linear (nullspace) components independently of the non-linear (range space) components (having different degrees of smoothing), while the alternate approach first removes the wiggly parts of the model, smoothing towards a straight line (or nullspace) then once wiggles are removed, the linear parts are removed, smoothing towards a zero effect. Either could be implemented with the models presented in SC/67b/EM03.

CWH reported on investigations of the question of possible decreases in body condition in Antarctic minke whales during the JARPA years using the focused information criterion (Claeskens and Hjort, 2008; Cunen *et al*., submitted). A total of five response variables were studied. Guided by previous comments in the scientific committee, the class of linear mixed models (LMM) was used, since these models made it possible to account for dependencies between the observations. The results showed that all five response variables had declined substantially over the JARPA years, and that all declines were statistically significant at the 5% level. For blubber at the BT11 position the wide model (which was considered to include all possibly meaningful covariates) had an estimated year effect of -0.0178 cm/year with standard error of 0.0071. As indicated in Table 3, the FIC model with optimal estimation capability had an estimated year effect of -0.0186 cm/year with standard error of 0.0066. The decline in blubber thickness was somewhat larger in the west than in the east, but no difference in decline was found between males and females, neither in the west nor in the east.

Table 1

Comparison of the variables used by McKinlay, de la Mare and Walsh (MDW) and Cunen, Walløe and Hjort (CWH) in their 2018 analyses.

Variable	MDW	CWH
Response variable	BWt (revised), BT11 (new)	FW, BT11, BT7, AG, HG
Year	Continuous (NP, 7 df BS) [NP: nonparametric, hereafter] [BS: before shrinkage]	Linear + quadratic
Sex	Separate analyses Female/Male	On
Age	Continuous (NP, 8 df BS)	Off
Body length	Body weight	Body length
Body weight (as covariate in models of BT11)	Continuous (NP, 8 df BS)	Supplementary analyses including it as a linear covariate
Fetus length	Continuous (NP, 8df BS)	Linear
Diatom	Two level factor, Low/High	Two level factor, Low/High
Date	Continuous (NP, 8 df BS)	Linear + quadratic
Region	Separate analyses West & East, combined for illustration	Factor; West/East/RossSea
Spatial Location	Separate longitude effects for each level of ice (NP, 20 df BS)	Latitude as a Linear covariate (Longitude only via Region)
Distance from ice	Two level factor, Near/Far from ice edge	Supplementary analyses with Ice as a factor –no effect
Interactions	Year x Date; Year x FetusLength; Year x Age; Date x FetusLength; Age x Length; Year x Length; Weight x FetusLength	Diatom x Date; Diatom x Date^2; Latitude x Date; Latitude x Date^2; Year x Region; Year^2 x Region; Latitude x Region; Sex x Region; Diatom x Region; (implicitly via random effects: Date x Year); (have checked: Year x Sex; Year^2 x Sex; Year x Sex x Region; Year^2 x Sex x Region)

Table 2

Comparison of inference used by McKinlay, de la Mare and Walsh (MDW) and Cunen, Walløe and Hjort (CWH) in their 2018 analyses.

Inference	MDW	CWH	Notes
Model	Generalized additive models utilizing routines in R package mgcv.	Linear mixed-effect models	
Contrasts used; implications	Treatment; results based on predictions that condition on treatment levels.	Sum-to-zero for FIC analyses, Treatment for presentation (in some cases). Two choices give the same fitted lines.	
Inferential method informing condition	Model predictions for early and late-season median length and median age animals, the difference of which provides a standardized relative index of the within season accumulation of weight or BT11.	REML. Look at the year effect directly: focus is the year coefficient (when there is only a linear year effect) or the mean of the derivatives of the response w.r.t year (when there is also a quadratic year effect).	This is the most critical, and significant difference between the two approaches

MDW argued that the small increase in whale lengths during the JARPA years could explain the decrease in blubber thickness. Because of this possibility, CWH considered total fat weight, which is the sum of the weight of subcutaneous fat (“blubber”) and the weight of the intestinal fat dissected out during the flensing of the whale to be a more reliable measure of storage of fat. This variable also showed a substantial and significant decline over the JARPA years when LMM were used. In the winning FIC model, with its selection of bias variance trade-off, there was an estimated year effect of -0.0073 tons/year with a standard error of 0.0023. Again, only small differences were found between the two sexes.

Table 3

Change in focus variable with years.

Variable*	Slope Estimate	SE	t	p-value#
Blubber at BT11	-0.0186 cm/year	0.0066	-2.83	0.0046
Fat Weight	-0.0073 tons/year	0.0023	-3.10	0.0019

* Similar result for the three other response variables. Only small difference between the two sexes.
#P-value for both sides test.

CWH noted that in SC/67b/EM01 and SC/67b/EM03 a new variable of primary interest was introduced by MDW, namely the 'accumulated blubber thickness in each feeding season'. CWH considered that the prediction process used in these papers for estimation of summer improvement in condition relied on a number of uncertain assumptions and that the summer accumulation of blubber thickness did not fully reflect the potential year trend. For instance, if one season had been "bad" the whales would probably start the next season at a lower level, while the summer accumulation would stay the same or even increase. As a result, CWH considered this choice of focus variable to be less informative than looking directly at the effect of year.

In discussion it was noted that marine mammals use both fat stores and protein stores for energy when fasting. Therefore fat stores are not the only location from which mass will be lost during fasting or periods of low food availability. A participant suggested that this was a sufficient reason to include total body weight in the models for Antarctic minke whale condition, especially as blubber mass varies with body size and given the individual plasticity in fasting physiology. An additional suggestion was made that further progress may be made on this issue by inviting a larger group of physiologists to take part in the discussions.

MDW provided Table 4, which gives the inverse variance weighted linear regressions of the yearly estimates of accumulated blubber thickness that were presented in SC/67b/EM03. The response was accumulated blubber thickness between the 20th percentile of sampling data for low diatom animals and the 80th percentile of sampling date for high diatom animals. Thus the analyses examined the gain in condition that can be attributed to summer feeding and accounting for the interactions between body length and weight, which were shown to be an important feature of the data. Full details of conditioning for predictions are in SC/67b/EM03.

Table 4
Inverse variance weighted linear regressions of the yearly estimates of accumulated blubber thickness.

Model	Year trend	SE	t	p-value
Females East	+0.0029	0.0099	+0.290	0.780
Males East	+0.0036	0.0101	+0.359	0.730
Females West	-0.0378	0.0004	-9.174	<0.001*
Males West	-0.0100	0.0142	-0.703	0.509
Combined	**-0.0097**	**0.0122**	**-0.789**	**0.436**

* The statistical significance is substantially overstated because the predictions come from a smoothed predictor that was nearly linear.
# P-value for both sides test.

In summary, MDW noted that there were no substantial trends in three of the four subdivisions by Sex and Region (East or West). Only for females in the West was there evidence of a decline. The combined results had a negative trend, which was not significant. However, MDW considered that claims for a negative overall trend were misleading because the year trends were only negative and significant for females in the Western half of the JARPA survey area. For the purpose of ecological modelling the authors consider that the regions and sexes should be treated separately.

There was general agreement between MDW and CWH that the choice of an initial model for any analysis should be as inclusive as possible, incorporating all relevant explanatory variables and interaction terms. However, MDW considered that the results of FIC analyses were dependent on the selection of the wide model. MDW considered that that an issue with the FIC method was that there appears to be no formal criteria for model selection of the wide model, presenting a difficulty given their belief that the results of the FIC selection were dependent on the wide model estimate of the 'focal parameter'.

The summaries of analyses by CWH and MDW are respectively given Appendices 2 and 3.

At the conclusion of the discussion, statements were made regarding the Data Availability Agreement related to the body condition analysis for the use of JARPA data. These statements can be found in Appendix 4.

2.1.3 Conclusions

The Scientific Committee agreed by consensus at its 2014 meeting that there had been a statistically significant (5% level) decline in blubber thickness and fat weight in Antarctic minke whales over the 18 years of the JARPA surveys [IWC, 2015]. Over subsequent years, multiple analyses have been presented to the Working Group supporting or arguing against this conclusion. Statistical estimation methods have been refined, and some analysts have (giving reasons, though these are not universally accepted) changed the variables which they consider for the evaluation of whether or not there has been a decline. The Working Group **agreed** that for the data set considered as a whole, all approaches result in point estimates reflecting a decline when fit to a linear trend in time. However, the extent of the decline estimated differs amongst the methods, and is not statistically significant at the 5% level for all approaches. Furthermore, for some approaches, when the data are disaggregated by gender and/or area, some point estimates of trend are not negative; in addition, there are some indications of temporal variation that is more complex than linear. The Working Group thanked the authors for their dedicated efforts towards refinement of their analyses and encouraged them to publish the results of their studies as soon as possible. The Working Group **agreed** that this matter need not be discussed further before the 2021 meeting at the earliest.

2.2 Review approaches used in body condition analyses for other stocks
The Chair welcomed the information that there would be a bowhead whale body condition analysis presented to the Working Group within the next two years. He also encouraged other members in the Working Group to bring forward relevant research.

3. REVIEW ISSUES RELEVANT TO ECOSYSTEM MODELLING WITHIN THE COMMITTEE

3.1 Individual-based energetic models
SC/67b/EM07 outlines enhancements to the individual-based energetics model (IBEM) developed since last meeting (also discussed in the RMP Sub-Committee, see Annex D). One of these changes enabled feeding on migration to be explicitly modelled. The model now also allows for more detailed foraging behaviour including the modelling of individual dives (de la Mare, submitted, SC/67b/FI28) and searching for prey schools (SC/67b/EM04). Results presented for 'minke like' whales showed that carrying capacity and productivity were sensitive to the level of food available during migration. An important implication is the need for ecosystem models to cover the entire migratory range of the species.

The Working Group noted that contribution of SC/67b/EM07 with regards to the determination of species functional responses is a valuable contribution for ecosystem modelling. The explicit foraging behaviour also enables the investigation of varying costs of foraging over different ranges of prey density and species abundance.

3.2 Modelling of relationship between whales and prey
De la Mare (submitted, SC/67b/FI28) provided an update on the individual based model of feeding diving behaviour. The model is process based, using high-resolution data from suction cup tags that record the characteristics of dives and individual feeding lunges. These data enable the calculation of functional relationships, which describe food consumption and energy gain as functions of the density of locally available food. Functional responses are central to the development of ecosystem models. The model demonstrated likely differences in the functional responses of two species, blue whales and minke whales. These responses indicated that blue whales were more efficient at exploiting prey at lower densities. The analyses also demonstrated that functional responses can depend on length of daylight and the vertical distribution of prey, particularly when prey density is measured by integrating over depth (e.g. $gm.m^{-2}$).

The authors emphasised that the purpose of this exercise was to illustrate the properties of the model and determine what might be done to improve it. One such improvement was the inclusion of more realistic prey fields based upon empirical data. However, it is an advance over previous attempts to estimate functional response.

SC/67b/EM04 described a model that investigates the role of searching for suitable food patches when developing functional responses. The model relates large-scale prey average density ($gm.m^2$) to the probability that an animal transiting in a region will detect a prey school within a given distance of its track. The prey model is based on Brierley and Cox (2015) which shows that changes in large-scale density tend to arise from changes in the number of prey aggregations rather than changes in their characteristics (such as volume and density). The resultant functions are sensitive to the distance at which whales are able to detect prey aggregations. However, in relation to the density of krill in the Antarctic the large scale densities found in surveys are such that the model predicts that the time whales spend searching for krill swarms is unlikely to be the most important effect in defining the functional response.

Different areas of active research aimed at measuring blue whale's detection distance to krill swarms were discussed. These included the use of drones and photogrammetry, and the use of focal follows to determine the distance between periods of area restricted search. However, it was noted that both approaches would be more challenging for krill swarms at depth.

Foraging grounds of the Antarctic blue whale were surveyed in the austral summer of 2015 during the joint New Zealand-Australia Antarctic Ecosystems Voyage. Using this dataset, SC/67b/EM06 described the distribution of these rare whales in relation to their main prey species, Antarctic krill. A combination of passive acoustic technology and visual observations were used to locate Antarctic blue whales, whilst simultaneously using active underwater acoustics to characterise the distribution and density of krill swarms. Results suggested that Antarctic blue whales were more likely to be present within the vicinity of krill swarms detected at night, those of higher internal density, greater vertical height, and those found shallower in the water column. This study demonstrated that using complementary, multidisciplinary technologies can provide insights into sub meso-scale (i.e. <100 km) foraging behaviour of rare whales in a challenging environment. The nature of krill aggregations preferred by Antarctic blue whales is an important consideration, not only for the management of this endangered species in a changing environment, but also for the management of Antarctic krill fisheries.

It was clarified that the time and distance resolution used in SC/67b/EM06 (1hr and 12km) were based upon observer distance. In the future, different scales will be explored to determine at what point the observed relationship no longer holds. However, a more detailed exploration of the data is required before this aspect of the analysis can be addressed.

3.3 Modelling of competition among baleen whales
It was noted that multi-species individual based energetic models (IBEM) could be used to model direct and indirect competition of different whale species in the same environment. A version of the program to model competition between humpback and minke whales in the Antarctic was nearing completion. The Working Group expressed interest in this work, and welcomed future submissions to the SC.

3.4 Effects of long-term environmental variability on whale populations
The Working Group noted that this was an active area of research, and was of particular interest to the SC with regards to how long-term environmental variability might affect stock assessments. The need for a literature review on the subject was highlighted, and the Working Group **agreed** to form an intersessional correspondence group with Cooke as convenor.

3.5 Stable isotope analyses
SC/67b/SP09 reported the preliminary results of a stable isotope analysis on samples from the edge of baleen plates in Antarctic minke whales. The aim of this exercise was to estimate the duration of the time whales had spent in the Antarctic feeding grounds. The stable carbon (δ13C) and nitrogen isotope ratios (δ15N) were determined from the edge of baleen plates of ten pregnant females in the Ross Sea, and six immature females sampled in the NEWREP-A surveys in 2016 and 2017. In the pregnant females, about four fluctuations of δ15N were seen at each baleen plate. The trophic enrichment factor (TE) was estimated to be 3.48%. This TE and δ15N values in all parts of baleen plate suggested that the whales fed mostly on Antarctic krill for a long period. In the immature animals, the temporal change of δ15N was high after birth, and was followed by a rapid decrease, probably indicating high values δ15N drops when they feed on krill. The growing ratio of baleen plates is the most important data needed to estimate the duration of time whales spend in the feeding grounds, and to determine the meaning of the fluctuations observed in the pregnant females. Knowledge of the behaviour of δ15N when the whales are fasting could also help in understanding the observed fluctuation. Geographical variations will be examined by analysing additional samples in the future.

4. ECOSYSTEM MODELLING IN THE ANTARCTIC OCEAN
The Working Group expressed interest in the continuation of plans for joint workshops with CCAMLR on ecosystem modelling in the Antarctic Ocean. It was **agreed** that a two-year delay in the occurrence of the workshop would provide the Working Group with the opportunity to pursue and complete the relevant work, and that in the interim they would seek information and advice from CCAMLR as needed. As in previous years, the Working Group **recommended** that collaboration between SC-IWC/SC CCAMLR be on going, and that the revised plan for the workshops be implemented by SC68b.

Attention: SC

The Working Group ***recommended*** *that collaboration between SC-IWC/SC CCAMLR be on going, and that the revised plan for the workshops be implemented.*

---

5. APPLICATION OF SPECIES DISTRIBUTION MODELS (SDMS) AND ENSEMBLE AVERAGING
An update on the intersessional correspondence group on the applications of species distribution models (SDMs) was presented to the Working Group. While there was not significant progress between meetings, the Working Group **agreed** that developing guidelines for best practice for species distribution models (SDM) was important, and that the correspondence group should be retained.

6. OTHER MATTERS

6.1 Review information on krill distribution and abundance by NEWREP-A
SC/67b/EM05 reported the results of the krill and oceanographic surveys during the third NEWREP-A survey in Area V-E and VI-W. The surveys, which are associated with the main objective II of NEWREP-A, were conducted by two research vessels *Yushin Maru* No. 2 (YS2) and *Kaiyo Maru* No. 7 (KY7). The surveys were conducted along the zig-zag tracklines designed for the whale sighting survey. Acoustic data using quantitative echosounders EK80 (YS2) and EK60 (KY7) were recorded continuously for total 73 days (6,608 nautical miles). Net sampling using small ring net (YS2 and KY7) and an Issak-Kid Midwater Trawl (IKMT) (KY7) was carried out to identify species and size composition of plankton echo signs at 47 stations and 11 stations, respectively. Oceanographic observations were also conducted at 112 stations using a Conductivity-Temperature-Depth profiler (CTD) and seawater sampling occurred at 16 stations. Calibration among EK80 and EK60 quantitative echosounders, and simultaneous samplings between small ring net and IKMT were also conducted. Krill and oceanographic data are currently being examined, and results will be reported to relevant CCAMLR working group and the mid-term review of NEWREP-A.

In discussion of SC/67b/EM05 it was noted that an objective of the NEWREP-A survey is to study the variation in abundance of different krill species in order to define the prey landscape as relevant to the whales. Given the difficultly in distinguishing species acoustically, the authors are currently using net sampling to ensure species identification and are pursing on-going research with experts to be able to identify species based solely upon the data collected by the echosounder. This information was welcomed, as it would be a large step forward in ecology and ecosystem modelling, but achieving this objective is likely very difficult.

It was also clarified that the departure from the expected krill survey design recommended by CCAMLR was due to the data being collected in conjunction with a survey for whales. In particular, there is simulation work underway to compare the bias and precision of the krill biomass estimates that might be achieved from the NEWREP-A survey design versus that which will be achieved following the standard survey design advocated by CCAMLR. In addition, data collection has been increased in an attempt to obtain the required number of samples. A concern was raised that data collected from EK60 and EK80 echosounders may not be comparable.

6.2 Ecosystem functioning

Resolution 2016-3 tasked the Working Group with investigating the contribution of cetaceans to ecosystem functions. The interessional correspondence group (ICG-28) reported progress in this area, developing terms of reference that included: (1) determining how best to integrate this task into the EM Working Group, (2) Reviewing and developing pathways between cetaceans and ecosystem services, as well as integrating these into ecosystem models, and (3) develop a gap analysis regarding research in the contribution of cetaceans to ecosystem models and identify needed research. The importance of identifying an area, or areas, where such research could take place was also highlighted. There is broad interest in understanding the role of cetaceans in ecosystem functions, as evidenced by the recent Convention on Migratory Species (CMS) and the formation of a Standing Working Group on Cetaceans and Ecosystem Functioning in the Conservation Committee, in which it is acknowledged that migrating cetaceans provide valuable contributions to ecosystem function. Specifically, the Working Group will be looking at the scientific aspects of ecosystem functions, while the Conservation Committee will focus on the conservation and social science aspects of the issue. Further, CMS has expressed interested in collaborating with the IWC to advance development on this issue.

In discussion the need to develop a framework to approach this complicated issue was raised. It was **agreed** that in order to understand the contribution of cetaceans to ecosystem services it was first necessary to identify broad gaps in our knowledge, the broad ecosystem function categories of interest, data needs and availability, and geographical locations as well as species that may be suitable for exploration of this issue. Only after this was achieved could detailed modelling approaches be considered, but first feasibility studies would need to be undertaken (see subsequent discussion at end of this section). It was determined that the Southern Ocean was the mostly likely place where the role of cetaceans in ecosystem function could be studied, because smaller regions would suffer from complications when accounting for immigration and emigration. Another advantage of the Southern Ocean was the possibility of cooperation with CCAMLR, and other organizations such as CSIRO, to share relevant data. Additional organisations, such as the Scientific Committee on Ocean Research (SCOR) and the Scientific Committee on Antarctic Research (SCAR), are already doing research in this area that is relevant, particularly with regards to ocean ecosystem variables.

Lastly, it was noted that even where data available and study systems well defined, it may not be possible to determine the contribution of cetaceans to ecosystem services. Many researchers have studied multi-species interactions and ecosystem models over the years, but general conclusions have been difficult to draw. The Working Group agreed that is was worthwhile to pursue research into the contribution of cetaceans to ecosystem functioning, but that it must be done with the acknowledgment that the work may not be as successful as anticipated and that it is unlikely that the ultimate goal of fully determining the contribution of cetaceans to ecosystem functioning could be achieved in under a decade. A more immediate and achievable goal is the carrying out of a gap analysis to address knowledge gaps. Therefore, the Working Group agreed that it would be beneficial to hold a workshop to define clear objectives and determine what further research is required in order to begin modelling the contribution of cetaceans to ecosystem function, and that CMS should be approached to determine their interest in participating in such a workshop.

7. WORK PLAN AND BUDGET REQUEST

7.1 Workplan for 2019-2020

The Working Group discussed the proposed details of the workshop agreed upon in Item 6.2, which would be integrated with the work being done by the Sub-Committee CMP. It was noted that there were many researchers already working on the understanding the contribution of species to ecosystem function, including ecosystem modellers. As a result, it was suggested that the list of workshop participants try to represent the different research groups with experience working in this area. The workshop would then provide an opportunity to bring people together and to combine resources to avoid the problem of different groups attempting to solve the same problem. The Working Group **endorsed** the workshop proposal.

In response to the proposed workshop, Japan made the following statement.

As Japan expressed at the occasion of the adoption of Agenda of this SC, it does not support the SC to deal with issues outside the competence of IWC. It found that a number of activities envisaged to be dealt with at the proposed workshop are outside the competence of IWC. For this reason, Japan cannot support the proposed workshop, and especially it cannot support the allocation of a budget of SC for this purpose.

Table 5

Summary of the two-year work plan for the EM working group

Item	Intersessional 2018/19	2019 Annual Meeting (SC/68a)	Intersessional 2019/20	2020 Annual meeting (SC/68b)
(1) Ecosystem modelling in the Antarctic Ocean	Continue further analyses.	Review results of further analyses	Continue further analyses.	Review results of further analyses
(2) Application of species distribution models (SDMs)	Intersessional Working Group activity (see Annex xx)	Review progress Working Group.		
(3) Effect of long-term environmental variability on whale populations	Continue further analyses. Intersessional Working Group activity (see Annex xx)	Review results of further analyses. Review progress Working Group.	Continue further analyses	Review results of further analyses
(4) Further investigation of individual-based energetic models	Continue further analyses	Review results of further analyses	Continue further analyses	Review results of further analyses
(5) Modelling of competition among whales	Continue further analyses	Review results of further analyses	Continue further analyses	Review results of further analyses
(6) Update of any exercises on krill distribution and abundance	Conduct NEWREP-A krill survey and an international cooperative krill survey. Conduct simulation analyses to resolve issues on survey design.	Review results of survey and analyses.	Conduct NEWREP-A krill survey. Conduct analysis of data taken by the international survey.	Review results of survey and analyses.
(7) Cetaceans & Ecosystem Functioning: a gap analysis workshop	Review relevant scientific studies before workshop (or premeeting) in addition to preparation of workshop.	Review outcomes of workshop and develop clear work plans with priorities.	Continue analyses	Review results of analyses.

Table 6

E-mail Intersessional Correspondence Groups, Steering Groups, Working Groups and Terms of Reference.

Group	Sub-committee	Terms of Reference	Membership
(1) Applications of species distribution models (SDMs)	EM	Develop guidelines and recommendations for best practice in species distribution modelling	Murase (convenor), Friedlaender McKinlay, Miller, Kelly, Kitakado, Palacios, Palka
(2) Effects of long-term environmental variability on whale populations	EM	Compile a literature review on the subject of how environmental variability may affect whale populations	Cooke (convenor), Butterworth, de la Mare, Kitakado
(3) Cetaceans & Ecosystem Functioning: a gap analysis workshop (Steering group)	EM	Prepare a Workshop under a Steering Group	Ritter (convenor), Butterworth, Donovan, Galletti, Kitakado, Suydam

7.2 Budget requests for 2019-2020

Table 7 summarises budget request for Ecosystem Modelling Working Group.

Table 7

Summary of the 2-year budget request for Ecosystem Modelling

RP no.	Title	2019 (£)	2020 (£)
	Meetings/Workshop		
EM-1	Cetaceans & Ecosystem Functioning: a gap analysis workshop	£20,300	
	Total request	£20,300	

8. ADOPTION OF REPORT

The report was adopted on 02 May 2018 at 12:22. The Chair expressed his sincere appreciation to the rapporteur, New, for her excellent work and thanked the participants for their valuable contributions. The Working Group thanked Kitakado for his leadership and gratefully accepted his offer to convene the Group next year.

REFERENCES

Brierley, A.S. and Cox, M. 2015. Fewer but not small schools in declining fish and krill populations. Current Biology, 25(1), 75-97. doi:10.1016/j.cub.2014.10.062

Claeskens, G. and Hjort, N.L. 2008. *Model Selection and Model Averaging.* Cambridge University Press.

Cunen, C., Walloe, L. and Hjort, N.L. In review. Focused model selection for linear mixed models with an application to whale ecology. *Annals of Applied Statistics*

Cunen, C., Walloe, L. and Hjort, N.L. 2017. Decline in energy storage in Antarctic minke whales during the JARPA period: Assessment via the Focused Information Criterion (FIC). IWC document SC/67A/EM04 presented at meeting 69th Annual Meeting of the Scientific Committee, Bled.

de la Mare, W., Friedlaender, A. and Goldbogen, J. Submitted. Developing function responses using an individual-based energetics model for rorqual foraging dives. *Functional Ecology*

McKinlay, J., de la Mare, W. and Welsh, A. 2017. A re-examination of minke whale body condition as reflected in the data. IWC document SC/67A/EM02 presented at meeting 69th Annual Meeting of the Scientific Committee, Bled

Appendix 1

AGENDA

1. Introductory items
    - 1.1 Convenor's opening remarks
    - 1.2 Election of Chair
    - 1.3 Appointment of rapporteurs
    - 1.4 Adoption of Agenda
    - 1.5 Documents available

2. Body condition analyses
    - 2.1 Review progress of analysis for Antarctic minke whales
        - 2.1.1 Review results of analyses
        - 2.1.2 Discussion
        - 2.1.3 Conclusion
    - 2.2 Review approaches used in body condition analyses for other stocks

3. Review issues relevant to ecosystem modelling within the Committee
    - 3.1 Individual-based energetic models
    - 3.2 Modelling the relationship between whales and prey
    - 3.3 Modelling of competition among baleen whales
    - 3.4 Effects of long-term variability on whale populations
    - 3.5 Stable isotope analyses

4. Ecosystem modelling in the Antarctic Ocean

5. Application of species distribution models (SDMs) and ensemble averaging

6. Other matters
    - 6.1 Review information on krill distribution and abundance by NEWREP-A
    - 6.2 Ecosystem functioning

7. Work plan
    - 7.1 Work plan for 2019-2020
    - 7.2 Budget requests for 2019-2020

8. Adoption of report

Appendix 2

BODY CONDITION ANALYSES IN ANTARCTIC MINKE WHALES

Céline Cunen, Lars Walløe and Nils Lid Hjort

Introduction

In Cunen, Walløe, and Hjort (2017) and in later contributions (for instance SC/67b/EM02 of this year), we have investigated the question of decreasing body condition in Antarctic minke whales during the JARPA years. We have devoted considerable energy into developing and motivating a biologically plausible model for the measurements of body condition (we call this 'the wide model'). We have looked at a total of five, correlated response variables. In this appendix, we summarize our analysis of blubber thickness (BT11), because McKinlay *et al* have analysed this dependent variable in their analyses (SC/67b/EM01 and SC/67b/EM03), but we have similar results for the four other responses. Guided by previous comments in the scientific committee, we have used the class of linear mixed models, since these models make it possible to account for dependencies between the observations. Specifically, we let several terms in our model be affected by random effects, which let observations from the same season (year) be correlated. The main question is whether there has been a considerable and statistically significant decrease in body condition, which in our regression framework translates into whether we can claim that the parameters describing the effect of year are negative and significant on any usual level of significance (when we have a linear year effect, there is only one parameter describing the year effect).

In large biological models like the one we have used here, there is often a need for some model selection criterion. The (main) reason is that the models that are biologically plausible and well-motivated, usually are very big (i.e. have many parameters) since we have many different predictor variables that may influence the response. Such a big model will often lead to large uncertainty around the parameter estimates and thus make it difficult to draw clear conclusions from data. Model selection aimed at simplifying a large model can be conducted in various ways, some ad hoc and others more principled. A common example of such a procedure is the usual backward selection of predictors; another example is in a way penalised regression methods. Here, we have used the focused information criterion (Claeskens and Hjort 2008), which offers a principled way of reducing the number of parameters in a large model **when one has a specific focus in mind**. The focus parameter is the parameter of main interest, and needs to have a clear statistical interpretation across candidate models. Here it is natural to focus on some measure of the year effect (we will come back to how we define this), since this informs the main question of the analysis.

Since the existing FIC framework did not cover the class of linear mixed effect models, we have developed a new FIC for the occasion (see Cunen, Walløe, and Hjort (2018)). We have used FIC to find a simpler model which describes the year effect as precisely as possible. This sentence must not be understood as FIC always finding a model with a significant focus parameter: if the estimated focus parameter is small compared to the variance, FIC can select a model where the focus parameter is zero. We have also analysed the wide model in itself. Both in the model selected by FIC (the winning model) and in the wide model we obtain negative and significant year effects. As expected (given the aim of FIC), the model selected by FIC estimates the year effect with better precision. Our finding is also strengthened by the fact that FIC gives a very bad score to any linear mixed effect model not containing the year effect: if the signal in the wide model had been weak, a model without year effect would have been preferred (since that model has zero variance), but that was not the case. In our view, there is therefore sufficient evidence that there has been a decline in body condition during the JARPA period.

Note that, here we use the same wide model as in SC/67b/EM02, but we have investigated several small alterations to this model. Specifically, we currently favour a model similar to the model in SC/67b/EM02, but with an extra second order interaction term $Sex * Year * Region$, which allows us to investigate differences in year effect between regions and males/females. Results from using this model have been reported on in EM08 and importantly, **they do not change the main conclusions we present here**.

Wide model and candidate models

We have used a similar wide model to the one in Cunen *et al*.(2017),

```
cc1 <- lmer(BT11~YearNum + I(YearNum^2) + BLm + Sex + DiatomF +DateNumS
        + I(DateNumS^2) + LatNum + Sex:Fetus.length + Sex*DiatomF
        + DiatomF*DateNumS + DiatomF*I(DateNumS^2) + LatNum*DateNumS
        + LatNum*I(DateNumS^2) + Region + YearNum*Region
        + I(YearNum^2)*Region + LatNum*Region + Sex*Region
        + DiatomF*Region + (1 +DateNumS + I(DateNumS^2)|YearNum),
        data=df1,contrasts = list(Region="contr.sum"))
```

The few alterations compared to the wide model in Cunen *et al*. (2017) are described in SC/67b/EM02. The wide model contains a second order effect of year. This allows the wide model to handle more complicated trends than simply linear ones. A natural definition of the focus parameter in the wide model is then $\mu = \beta_{year} + 2\beta_{year2}\overline{x}_{year}$, where the term $\overline{x}_{year}$ is the mean year in the dataset and β_{year2} is the coefficient corresponding to the second order year term. The focus parameter corresponds to the derivative of the response with respect to year, and then evaluated in the mid-point year; it may also be interpreted as the mean of the derivative values of the response, across the years of study. Crucially, since the candidate models are supposed to be as simple as possible, and since the second order term appears to be quite small, we have only included candidate models with linear year effect. For these the focus parameter is simply $\mu = \beta_{year}$ as before.

The model defined above has $p = 27$ fixed effect coefficients. The notation $(1 + Date + Date^2|Year)$ specifies the random effect structure; the groups are defined by a categorical version of the year variable. We have 3 random effects giving a total of 34 parameters to estimate. For now, we have limited ourselves to investigating 22 candidate models (check SC/67b/EM02 for model specifications). The models differ in their random effect structure, their inclusion of $Region$, the inclusion of some interaction terms, and a few fixed effects. The last model, M_{22}, is a baseline model without any year effect, so $\mu_{M_{22}} = 0$.

Interpreting FIC results

The focused information criterion ranks the different candidate models according to how precisely they estimate the parameter of main interest (the focus). Precision is measured as mean squared error (MSE) under the wide model. The model with the lowest FIC score is the winning model according to FIC. In some cases, this could be the wide model itself, but usually it will be a smaller model. If the signal in the focus parameter is weak in the wide model, FIC will select a model not containing the focus parameter ($\boldsymbol{M_{22}}$ in our case): FIC thereby conducts an implicit test of the focus.

The FIC approach we have used in Cunen *et al*. (2017) and Cunen *et al*. (2018) hinges on a biologically plausible wide model (which the user must believe is wide enough to accurately represent the real data generating mechanisms) and consists of a principled way of reducing this wide model in order to obtain more precise estimates of a certain parameter of interest. Unlike the wide model, the candidate models need not necessarily be biologically plausible: the candidate models are primarily meant to produce good estimators and are not necessarily meant to be interpreted as being close to the true data-generating mechanism. Thus the candidate models need only contain covariates that influence the estimation of the focus parameter. The candidate models may therefore lack some covariates that are important in themselves, but do not influence the focus. When a candidate model is selected by the FIC score it must be interpreted in the following way: this candidate model accurately represents the focus parameter in the wide model, but the user still believes in the wide, not the candidate.

Results from model selection with FIC

We have 4,718 observations. The results of model selection are displayed in the FIC plot below. There, one can read of the root-FIC scores and estimates of the models. The scale in use for both $\hat{\beta}$ estimates and root-FIC scores is in centimeters. For ease of presentation, the figure is zoomed in on the best models. The winning model is M_6,

```
cc6 <- lmer(BT11~YearNum + Sex +DateNumS + I(DateNumS^2) + LatNum
        + Region + (1 +DateNumS + I(DateNumS^2)|YearNum),
        data=df1,contrasts = list(Region="contr.sum"))
```

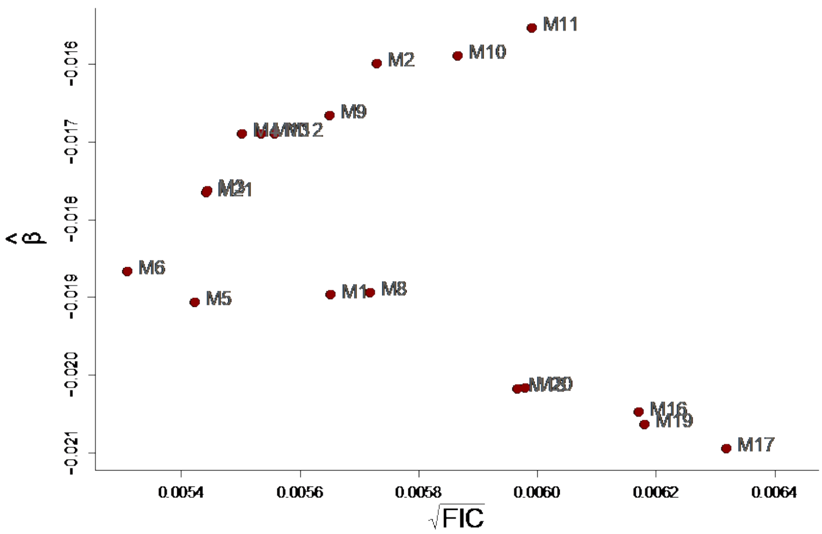


Fig. 1. FIC plot for BT11, showing only the best models. The baseline model M_{22} has root-FIC score of 0.016 and the wide model M_0 has root-FIC score of 0.0071 and gives a focus parameter estimate of -0.018.

Presenting the wide model and winning model

In the wide model we had an estimated year effect of **-0.0178** with standard error of **0.0071**; in the winning model we had an estimated year effect of **-0.0186** with standard error of **0.0066**. Note that this standard error is calculated **under** the wide model, i.e. we still believe in the wide model, but we use the winning model as a lower dimensional representation of the wide model. We can look at the full confidence curve for the focus parameter in the wide model. We see that it is significant at most common levels of significance. We can also compute the confidence curve for the winning model M_6. Again, the confidence curve is computed **under** the winning model. We see that it is narrower and has a slight bias compared to the wide model.

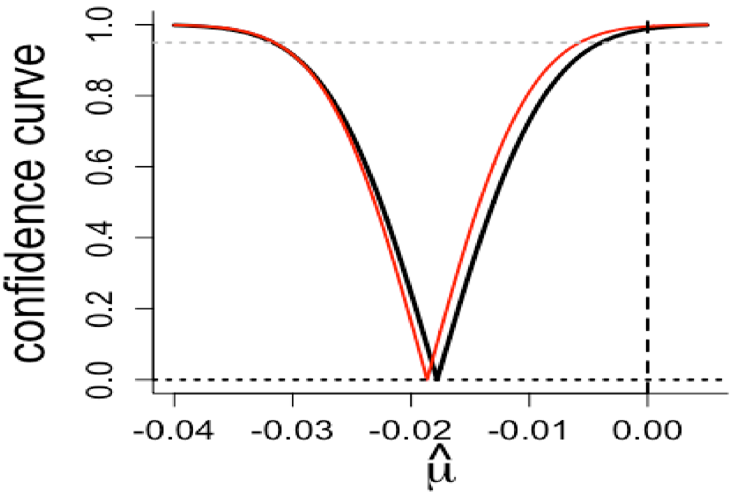


Fig. 2. Confidence curves for the effect of year, for the wide model (black) and the winning model (red).

Problems with MDW

In *No substantial change in Antarctic minke whale condition during the JARPA years* (McKinlay, de la Mare and Welsh, 2018), the authors, hereafter referred to as MDW, model two responses, body weight (BWt) and blubber thickness (BT11), using the framework of generalized additive models (GAMs). They split the available data into four parts according to sex and region (west/east), and provide some arguments for this choice. In their presentation of the results, they introduce a new focus parameter (a variable of primary interest), namely the *accumulated blubber thickness in each year*. We will focus our attention on the BT11 analyses in McKinlay *et al*. (2018). The choice of response variable was discussed in great lengths in the 2017 Scientific Committee meeting and in our opinion, several delegates agreed that total body weight constitutes a less relevant measure of body condition compared to the available measurements of blubber thickness, fat weight and girth. The following is a summary on our assessment of MDWs contribution (full text in SC/67b/EM08).

1. We consider MDW's choice of focus (the accumulated blubber thickness in each year) to be less informative than the more natural choice of looking at the effect of year directly. The accumulated blubber thickness does not fully reflect the potential year trend. For instance, if one season has been "bad" the whales will probably start the next season at a lower level, while the accumulation stays the same or even increases. Also, we do not believe that MDW have chosen a suitable model for assessing the accumulated blubber thickness: the model should have included the year covariate only through its interaction terms and not as a main effect. In addition, we are skeptical to the way MDW define their early and late season whales – we are concerned that some individuals in poor condition may have been inadvertently excluded from the predictions, it seems like they might condition on a late season whale being healthy, which will likely obscure any patterns of decreasing body condition. Especially for the size of fetuses late in the feeding season, is possible that females with small fetuses have experienced bad feeding conditions resulting in a delayed estrus and/or a slow growth of the fetus. The result could well have been that McKinlay *et al* select away from their analyses whales that have experienced bad feeding conditions even if they have been in the Antarctica the whole austral summer.
2. It is our opinion, when looking at the partial effects plots of year effect in MDW's models, that their results are broadly consistent with ours, and with our opinion that there has been an overall decrease in body condition. Their models show that for most groups of whales there has been a decrease in body condition over the JARPA years. For some groups the decreased has been more pronounced than the overall decrease, while for other groups it has been less pronounced. The reason MDW reach opposite conclusions compared to ours in SC/67b/EM03 (McKinlay *et al*. 2018), lies not in our (mostly) linear approach, *nor in "our* failure to" include body weight and length as the size controlling variables The differences in conclusion are primarily due to MDW's non-standard and unusual choice of focus, namely the accumulated blubber thickness in each season. Although MDW have provided some arguments for this choice of focus parameter, we note that their analyzes are heavily influenced by the predictor values they choose to condition on.

3. Splitting up the data into four parts is unnecessary. On the contrary, we believe that one should allow the possibility of borrowing strength between regions, while also including interaction terms so that potential differences are taken care of. MDW claim that space and time are “exactly confounded”, but in our opinion this would only be correct if one believed in absolutely no smoothness in space and time. Splitting the data can make issues with uneven sampling worse, as it hinders the model to benefit from existing similarities between effects in the different groups defined by MDW.
4. We also find that the decline in blubber thickness is somewhat larger in the west than in the east, but we do not find any difference in decline between males and females, neither in the west nor in the east. MDW argue that the small increase in whale lengths during the JARPA years could explain the decrease in blubber thickness. Because of this possibility we consider the total fat weight, which is the sum of the weight of subcutaneous fat (“blubber”) and the weight of the intestinal fat dissected out during the flensing of the whale as a more reliable measure of storage of fat. This variable also shows a substantial and significant decline over the JARPA years. In the wide model we had an estimated year effect on fat weight of **-0.0073** with standard error of **0.0029**; in the winning model it had an estimated year effect of **-0.0073** with standard error of **0.0023**. Again only small differences between the two sexes.

REFERENCES

Claeskens, G. and Hjort, N.L. 2008. Model Selection and Model Averaging. Cambridge University Press.

Cunen, C., Walloe, L. and Hjort, N.L. 2017. Decline in Energy Storage in Antarctic Minke Whales During the Jarpa Period: Assessment via the Focused Information Criterion (Fic). SC/67a/EM04, 1–55.

Cunen, C., Walloe, L. and Hjort. N.L. 2018. Focused Model Selection for Linear Mixed Models, with an Application to Whale Ecology. *Submitted,* 1–38.

de la Mare, W., McKinlay, J. and Welsh, J. 2017. Analyses of the JARPA Antarctic Minke Whale Fat Weight Data Set. SC/67a/EM01, 1–57.

McKinlay, J., de la Mare, W. and Welsh, A. 2017. A Re-Examination of Minke Whale Body Condition as Reflected in the Data. SC/67a/EM02, 1–108.

McKinlay, J., W. de la Mare and A. Welsh. 2018. No Substantial Change in Antarctic Minke Whale Condition During the JARPA Years. 1–134.

Appendix 3

BODY CONDITION ANALYSES IN ANTARCTIC MINKE WHALES

John McKinlay and William de la Mare

1. Regressions of yearly estimates from model outputs for GAM models of minke accumulated blubber thickness
The Working Group suggested it may prove useful to examine linear regressions of the yearly estimates of accumulated blubber thickness that were presented in SC/67b/EM/03. The intent of this request was to put the two techniques, linear mixed effects (LME) models by Cunen, Walloe and Hjort (CWH) and generalised additive models (GAMs) by McKinlay, de la Mare and Welsh (MDW), onto an "even footing" for comparative purposes. We do not think the model outputs of CWH are entirely comparable with our own, but these regressions do facilitate an easy comparison of our own results, and they additionally reveal why CWH, fitting a global trend, might report a linear decline in condition.

The regression response is predicted accumulated blubber thickness between the 20th percentile of sampling data for low diatom animals and the 80th percentile of sampling date for high diatom animals. Full details of conditioning for predictions are in SC/67b/EM/03. Body weight (in addition to body length) is included as a covariate in GAM models to account for several features of these data, described shortly.

Fig. 1 shows a linear fit imposed upon the non-linearly derived yearly mean estimates from our preferred GAMs. Importantly, the GAMs included a main effect of body weight and a year*body length interaction. These results show that males and females in the East region have slightly increasing body condition over the period, while males and females in the West have declining condition. We note two important features: (i) of these trends, only the female decline in the West is significant at 5% (Table 1); and, (ii) the fit for males in the West, even though non-significant, is heavily influenced by two high leverage points in 2002 and 2004.

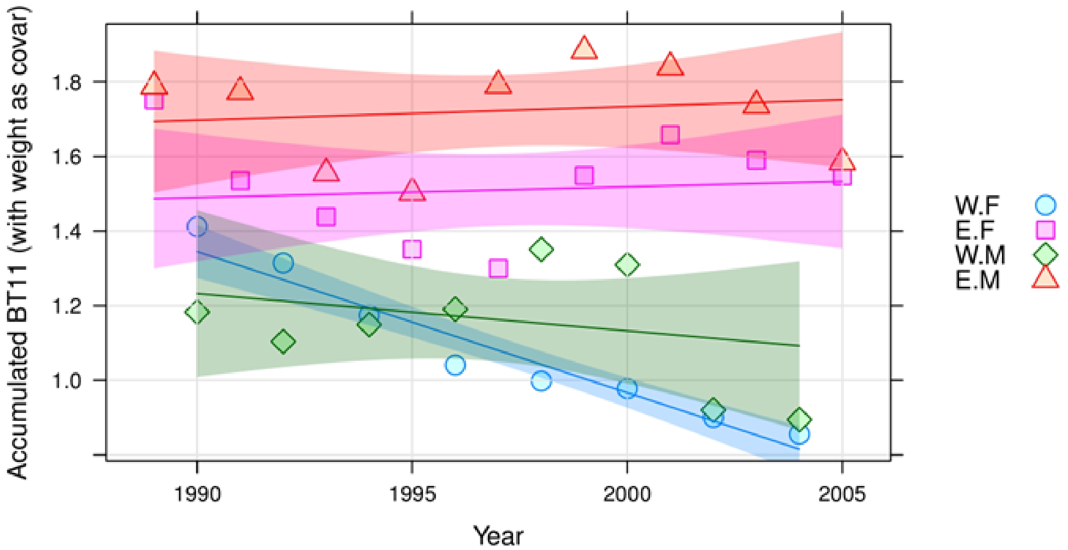


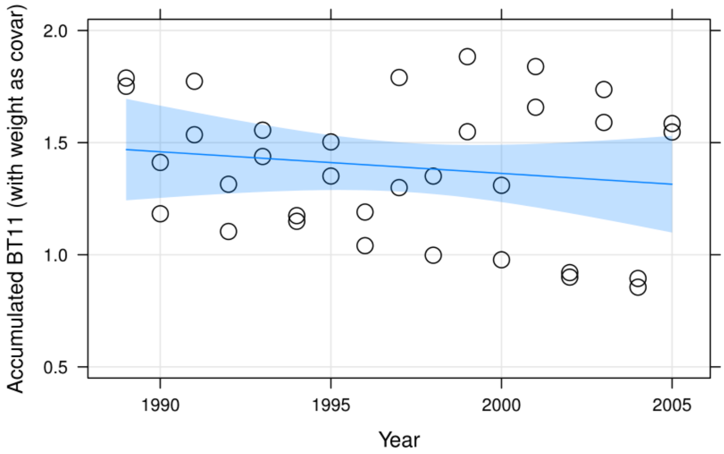
Fig 1. Regressions of year estimates of accumulated BT11 from GAMS, groups separate (males and females, West and East), including weight as an additional covariate in GAM models. Points weighted by $1/SE^2$ from GAM fits.

Table 3.1

Summary of slope estimates from linear regressions of yearly predicted accumulated blubber from GAM models that included body weight.

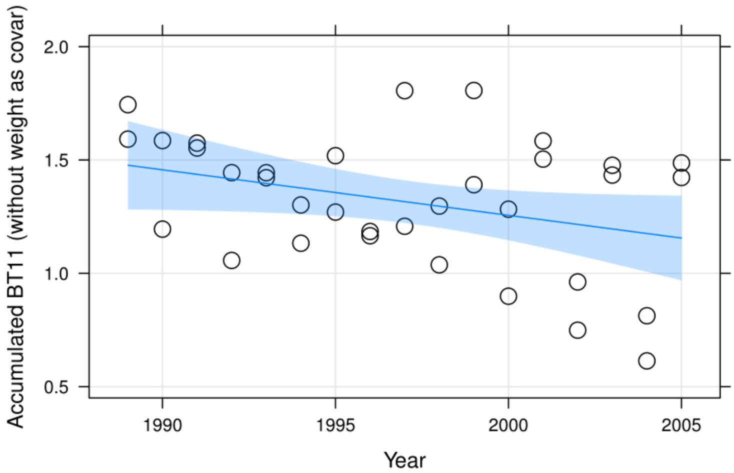
Model	Year trend	SE	t	P(>t)
West, Females	-0.0378	0.0004	-9.174	<0.001
East, Females	+0.0029	0.0099	+0.290	0.780
West, Males	-0.0100	0.0142	-0.703	0.509
East, Males	+0.0036	0.0101	+0.359	0.730

(a)



Slope	**-0.0097**
SE	0.0122
t	-0.789
P(>t)	**0.436**

(b)



Slope	**-0.0201**
SE	0.0105
t	-1.895
P(>t)	**0.067**

Fig. 2. Regressions of all groups combined (males and females from West and East) for year estimates of accumulated BT11 from GAMS: (a) including weight as an additional covariate in GAM models; and, (b) excluding weight as a covariate. Points weighted by $1/SE^2$ from GAM fits.

We next examine a linear fit to all the predicted means, ignoring the group structure (i.e. both regions and both sexes combined). This shows that the linear fit to the predicted means from the GAM that includes weight as a covariate has a slope very close to zero -0.0097 (SE 0.0122) and is non-significant at conventional levels (p=0.436) (Figure 2a). In contrast, the linear fit to the GAM results based on a fit ignoring weight as a covariate shows a much stronger linear decline of -0.0201 (SE 0.0105), and this is approaching significance at 5% (p=0.067).

We take two main points from these analyses: (i) you can get a misleading result if you take a single, strong trend (e.g. females in the West) and average it over several other weak or non-existent trends; and, (ii) including body weight in our GAM models has the effect of removing some apparent signal for declining condition (i.e. the slope in Fig 2a is smaller than in Fig 2b). We now show why this latter effect is occurring.

First, consider some simple linear models fitted to the raw JARPA data, showing blubber thickness (BT11) as a proportion of body length (cm/m) against body length (m) (Fig. 3). We show 95% CI for the fit, and exclude the data (because there is a lot of it). This clearly shows that *blubber thickness proportionately decreases with increasing body length*.

Next, consider the average catch lengths in the JARPA data, plotted against year (Figure 4). This clearly shows that, except for males in the East sampling region, *catch lengths have been increasing over time*. We do not know why this might have occurred in the Japanese Special Permit program, but it has.

We note that these two results will interact with one-another. If catch lengths go up, and blubber decreases with increasing body length, then this will provide a signal for decreasing blubber thickness unless these relationships are correctly captured in models. Including body weight in models achieves this goal, a point we demonstrate in SC/67b/EM/03 (Section 3.3, Figs 8-10).

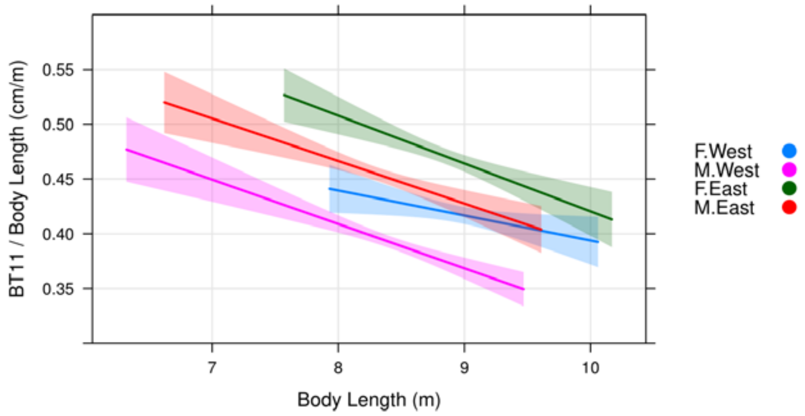


Fig 3. Blubber thickness (BT11) as a proportion of body length regressed against body length, showing proportionate decrease in blubber thickness with increasing length.

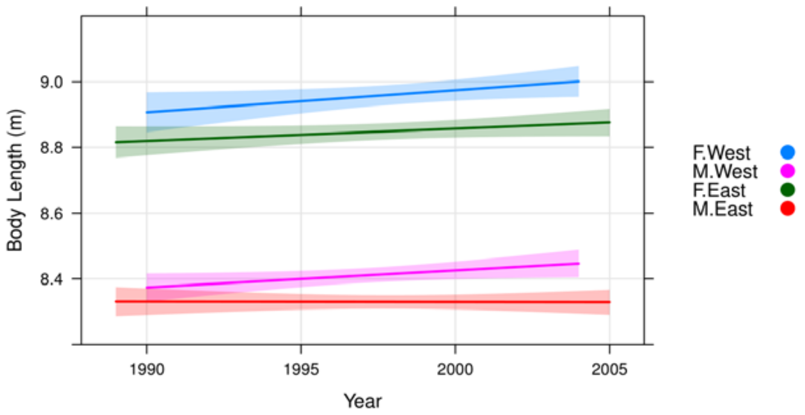


Fig. 4. JARPA catch lengths regressed against year for each group (sex and region).

2. Responding to a request from Norwegian scientists on how we think their models should be modified

During the discussions of the WG our Norwegian colleagues asked how they could accommodate into their models some of the issues we have discussed in our primary paper, SC/67b/EM/03. It is beyond the scope of this Appendix to do that topic justice and we refer the interested reader to our primary paper where we discuss the issues in detail. Further, as we do not subscribe to the philosophy of the Focused Information Criteria as a suitable model selection process, at least not in the case under consideration here, we are reluctant to try and suggest specific models. However, what we can point to are some important effects that we think should be captured in their models:

a. Diatom score is an important indicator of time spent feeding, so in our models we have tried to capture the **contrast** between low score animals captured at the beginning of the sampling season, and high score animals captured at the end of a season.
b. Average catch lengths in JARPA data have increased through time, a point easily demonstrated from the data (Fig. 4 above). Models should therefore capture this feature; we would suggest a year*length interaction.
c. Blubber thickness proportionately decreases with increasing body length; again, a feature that is easily confirmed from the data (Fig. 3 above). Given that catch length have been increasing over time, and that blubber decreases with length, failing to account for these features will induce a false signal for decreasing condition, or will exaggerate any real decrease that may be occurring. We have found that including both body length and body weight as covariates in models of BT11 has been the most effective way of capturing these aspects of the blubber-length-weight relationship. Assessing trends in blubber thickness when weight is included as a covariate in models requires careful conditioning.
d. Most importantly, in Fig. 1 we show that trends are different between the regions and sexes, and this should be modelled. We see no utility in fitting and reporting a single, overall trend. We do not think such an approaches is consistent with the stated requirements of the Scientific Committee, as reflected in the EM report from SC/67a.

3. Including total body weight as an independent variable is justified

Analyses of JARPA condition data need to correct for the effects of variations in the sizes of animals because, for example, the weight of fat an animal has depends on how large it is. At last year's EM meeting and again at this meeting there have been reservations expressed about using total weight as an independent variable as a measure of size in statistical analyses of body condition from JARPA data. The concern is that changes in total weight include the consequences of variations in other measures of body condition, such as fat weight. For example, because total weight includes fat weight, including the former as an explanatory variable leads to the claim that total weight cannot be considered as a valid independent variable. To examine this concern further we set up a simple model to determine whether this concern is consequential in the circumstances roughly applicable to the analyses of minke whale fat weight.

Whale blubber consists of several tissue types, some of which are primarily lipid stores and others have structural functions. Assume for simplicity that blubber weight is composed of a fixed structural component related to body size and a variable component reflecting energy storage. The total weight *W* of an animal is given by:

$$W = W_l + F_s + F_e \qquad (0.1)$$

W_l is the lean weight
F_s is the weight of structural blubber,
F_e is the weight of stored lipid (includes visceral fat as well)

Assume that the weight of structural blubber is directly proportional to lean weight, that is:

$$F_s = \alpha W_l \qquad (0.2)$$

Now suppose that there is a trend in energy storage (the central question) given by:

$$F_{e,t} = F_{e,1} + \beta t + \varepsilon_e \qquad (0.3)$$

where t is the year of sampling and ε_e is a normal random variable with mean = 0.0 and standard deviation σ_e . However, due the difficulty of collecting random samples further suppose that the sizes of animals, reflected in their lean weight, has a trend over time as well so that:

$$W_{l,t} = W_{l,1} + \gamma t + \varepsilon_l \qquad (0.4)$$

where ε_l is a normal random variable with mean = 0.0 and standard deviation σ_l . The only quantities that are directly observed are the total weight W_t and the total fat weight ($F_t = F_{s,t} + F_{e,t}$). Although lean weight can be calculated by subtracting the fat weight from the total weight, that does not eliminate the issue that the dependent variable (fat weight) is a component of an independent variable.

To anchor the model in the roughly the same numeric values as minke whales put:

$$\begin{aligned} W_{l,1} &= 8 \\ \sigma_l &= 0.1 \\ \alpha &= 0.125 \\ F_{e,1} &= 0.5 \\ \sigma_e &= 0.001 \\ t &= 1 \cdots 18 \end{aligned} \qquad (0.5)$$

Fitting a linear model to a single realisations of the model (with negligible random errors) gives:

Table 3.2

Model	Trend in W_l	True trend in F_e	Estimated trend in F	Bias
$F \sim t$	0.05	-0.015	-0.0087	-0.0063
$F \sim t + W$			-0.0128	-0.0022
$F \sim t$	-0.05	0.0	-0.0062	-0.0062
$F \sim t + W$			-0.0010	-0.0010

Including total weight (*W*) in the model has led to a considerable reduction in bias compared with using year (*t*) alone.

Appendix 4A

STATEMENT OF JAPAN REGARDING DATA AVAILABILITY AGREEMENT FOR BODY CONDITION ANALYSES

The Data Availability Agreement related to the body condition analysis for the use of JARPA data allowed the relevant parties to use them until the 2017 SC. In this regard all concerned issues were needed to be raised in 2017 SC, and the Norwegian scientists followed this condition and submitted comprehensive results to the 2017 SC and submitted a revised document to the 2018 SC in response to last year's comments, while the Australian scientists did not follow the agreed timeline.

Nevertheless, Japan allowed them to submit a document (SC/67b/EM03) to the 2018 SC, which included new analyses that were not based on last year's comments, in order to maintain collaboration in the SC. However SC/67b/EM01 rev submitted by Australian scientists only expands criticising other parties with new arguments which are not presented or even discussed at the 2017 SC. The collaborative work between the Australian delegation and the other two delegations failed due to such improper use of data by the Australian delegate.

The aforementioned DAA should be totally terminated at this SC, and the use of JARPA data treated under the ICR protocol.

Appendix 4B

RESPONSE FROM DE LA MARE AND MC KINLAY

De la Mare and McKinlay responded that their research plan and data request specified:

Although it is possible that work may be complete by the 2017 SC meeting, analyses are computer intensive and may need to be presented to the SC in a staged process. Final results, or at minimum a progress report, will be presented at the SC meeting in 2017.

We appreciate the pragmatic approach by the ICR, which has enabled us to bring this long standing issue to a conclusion at this meeting. However, we reject any implication that we have behaved improperly. Correspondence between us and the ICR demonstrates that we behaved ethically and, in particular, would not submit new analyses without ICR prior approval, which was provided to us. Criticism of research is a normal part of the scientific process.