Annex K1

Report of the Working Group on Ecosystem Modelling

Members: Palacios (Convenor), Atkinson, DeMaster, Baba, Balance, Bando, Baumann-Pickering, Bell, Bengtson Nash, Bjørge, Boyd, Brierley, Brownell, Butterworth, Charrassin, Currey, de la Mare, de Moor, DeAngelis, Diallo, Donovan, Double, Eisert, Elvarsson, Feindt-Herr, Fiedler, Fleming, Friedlaender, Funahashi, Gerber, Griffiths, Hakamada, Haug, Iñíguez, Johnson, Kelly, Ketele, Kitakado, Leaper, Lundquist, Manley, Mate, McKinlay, Mduduzi Seakamela, Melcón, Mesnick, Mikhno, Mizroch, Monnahan, Moore, Moronuki, Muraki, Murase, Natoli, Øien, Okazoe, Oosthuizen, Palka, Paniego, Park, Pastene, Perrin, Punt, Redfern, Reeves, Rendell, Reyes, Ridoux, Roel, Santos, Sigourney, Simeone, Širović, Sitar, Skaug, Solvang, Stachowitsch, Tamura, Thuok, Uoya, Velasco, Víkingsson, Wade, Walløe, Wang, Weinrich, Williams, Yasokawa, Zeh, Zerbini.

1. INTRODUCTORY ITEMS

1.1 Opening remarks
Palacios welcomed the members of the Ecosystem Modelling Working Group (hereafter, Working Group) and thanked the members of the intersessional E-Mail Steering Group (Butterworth, de la Mare, Punt, Walløe) for their guidance in planning the agenda for this year.

1.2 Arrangements for the meeting
The Working Group was allotted eight sessions this year to conduct its agenda. There would be a joint session with the Standing Working Group on Environmental Concerns. Topics that overlap with other sub-committees (such as RMP) would be handled within those sub-committees.

1.3 Election of Chair
Palacios was elected as Chair. He thanked the Working Group for their vote of confidence and promised to do his best to carry out the agenda.

1.4 Appointment of rapporteurs
Bell, Elvarsson and Kelly were nominated as rapporteurs. The Working Group approved of them unanimously and Palacios thanked them for their willingness to undertake this task. Butterworth would also assist in rapporteuring specific sessions.

1.5 Adoption of Agenda
Palacios proposed slight revisions to some of the standing items in the draft agenda to reflect recent and upcoming developments within the Working Group. The adopted Agenda is included as Appendix 1.

1.6 Documents available
The documents presented to and considered by the Working Group were SC/66a/EM01-EM05, Konishi and Walløe (In press), SC/66a/Rep07, SC/66a/Rep10, and IWC/66/4(2015) Annex C.

2. REVIEW OF ECOSYSTEM MODELLING EFFORTS UNDERTAKEN OUTSIDE THE IWC

2.1 Update from CCAMLR's Ecosystem Monitoring and Management Programme (WG-EMM) on krill and its dependent predators
Historically, the Working Group had reviewed relevant matters to ecosystem modelling from the IWC Observer's Report from the Annual Meeting of the Scientific Committee for the Conservation of Antarctic Marine Living Resources (SC-CCAMLR) (see IWC/66/4(2015)). Last year the Committee had appointed Currey as Observer to SC-CCAMLR and this year he highlighted the items in the 'Observer's Report from the 33rd Meeting of the Scientific Committee for the Conservation of Antarctic Marine Living Resources' (Annex C of IWC/66/4(2015)) relevant to the Working Group as well as more broadly to Committee as follows:

(1) a proposed joint CCAMLR-IWC Workshop on the development and application of multi-species models to the Antarctic marine ecosystem;
(2) coordination of photo-identification libraries;
(3) fish losses (primarily *Dissostichus* spp.) due to depredation by cetaceans, in particular killer whales (*Orcinus orca*) and sperm whales (*Physeter macrocephalus*);
(4) ecosystem interactions, particularly in relation to Type C killer whales in the Ross Sea (Eisert *et al.*, 2014a; 2014b); and
(5) baleen whale sightings associated with surveys between 2010 and 2014 near the South Orkney Islands (Krafft *et al.*, 2014; Orgeira *et al.*, 2014).

The Working Group welcomed these highlights. The subsequent discussion focused on the planning for a joint CCAMLR-IWC Workshop, and this is included under the next item.

Last year the Committee had also appointed Watters as Observer to CCAMLR's Ecosystem Monitoring and Management Programme (WG-EMM) on krill and its dependent predators. Watters could not be present at this year's meeting.

2.2 Update on planning for joint IWC-CCAMLR activities in 2016 and beyond
For some time the Working Group has had an ongoing interest in CCAMLR's activities, particularly those of its WG-EMM. In 2008, the IWC Scientific Committee had held a joint workshop with CCAMLR to review input data for Antarctic marine ecosystem models (IWC and CCAMLR, 2010). Since then, the Committee had identified significant knowledge gaps in aspects such as spatial variability and trends in prey species, on the relationships between predators and prey, and on the effects of environmental variability on predators.

In 2013, the Working Group had recommended consultation with CCAMLR through the formal channels of the Committee to establish collaboration between this Working Group (IWC/SC/EM) and SC-CCAMLR's WG-EMM (IWC, 2014b, p.333). Intersessionally during 2013/14, SC-CCAMLR had extended a formal invitation for an IWC

Scientific Committee observer to attend the annual meeting of the SC-CCAMLR WG-EMM. Last year the Committee had appointed Watters as Observer to SC-CCAMLR WG-EMM, and it had also appointed Currey as Observer to SC-CCAMLR. An intersessional Correspondence Group led by Currey had been established to continue formalisation of the relationship between IWC and CCAMLR and to commence planning of a joint Workshop to foster collaboration between IWC SC and SC-CCAMLR, including the development and application of multi-species models to the Antarctic marine ecosystem, as well as other activities that would be of mutual interest.

Currey provided the Working Group with an update on the proposal to hold a joint CCAMLR-IWC Workshop in June 2016, immediately prior to the SC/66b at the same venue (Bled, Slovenia). Discussion from the intersessional Correspondence Group established at SC/65b was presented at CCAMLR-XXXIII in October 2014. Following discussion of the proposal, CCAMLR-XXXIII recommended that the objectives of the joint Workshop be broadened to include each organisation as a whole, rather than just the Ecosystem Modelling Working Groups of each organisation. CCAMLR-XXXIII endorsed the formation of a Steering Group to progress the proposal. A preliminary list of SC-CCAMLR members was drawn up to work alongside representatives of the Working Group (see Item 5, work plan).

In discussion, this Working Group **agreed** to broaden the proposed objectives of the joint Workshop recognising the advantages of increasing facilitation of the collection and/or collation of data from within the scope of both organisations. However, the Working Group **recognised** the importance of maintaining a focus on the development and application of multi-species models to the Antarctic marine ecosystem. It **agreed** that the Steering Group would develop terms of reference for the joint Workshop intersessionally and consider participation and contributions to the Workshop against these (see Item 5, the work plan).

The Working Group encouraged the use of the joint Workshop as an opportunity to increase the collective state of knowledge on specific species and/or management areas (i.e. the six IWC management areas and three CCAMLR management areas, and identified overlap thereof). It was noted that a spatial focus may be necessary. The Working Group cautioned that some areas from which data are available were not representative of the broader Antarctic ecosystem (e.g. the Ross Sea). In contrast, the Working Group considered the Antarctic Peninsula to be more representative, as it is an area where substantial data already exist and where the potential for further data collection data is high (e.g. the USA's National Science Foundation Long-Term Ecological Research Project, CCAMLR's activities, the planned series of IWC-SORP R/V *Tango* voyages, and other IWC-SORP activities). It was also noted that the Antarctic Peninsula is high-priority area for CCAMLR in terms of achieving its objectives for the management of krill fisheries.

The Steering Group will address the spatial focus to adopt based on the data and participation determined to be available. Data requests will be submitted to the respective Secretariats of IWC and CCAMLR, particularly with regard to data resulting from the CCAMLR 2000 surveys (Reilly *et al.*, 2004; Watkins *et al.*, 2004). The Steering Group also should consider the ongoing efforts by this Working Group in the compilation of long time series of cetacean and environmental variables (see Item 4.2.1), as these data may be relevant to the objectives of the Workshop.

3. EXPLORE HOW ECOSYSTEM MODELS CONTRIBUTE TO DEVELOPING SCENARIOS FOR SIMULATION TESTING OF THE RMP

De la Mare (SC/66a/EM02) reported progress on using the individual-based energetics model (IBEM) in the exploration of the relationship between $MSYR_{1+}$ and $MSYR_{mat}$. This work is now primarily directed towards the work of the RMP Sub-committee (see Annex D, item 5.1 for further discussion). De la Mare also reported that the IBEM has been successfully incorporated into the RMP simulation software and some comparative results on some of the base-case trials will be reported next year. The natural extension of this work will be to include competition modelling (see Item 4.3) using the IBEM in the RMP testing framework. The next development of the IBEM will be tuned to minke whales.

4. REVIEW OTHER ISSUES RELEVANT TO ECOSYSTEM MODELLING WITHIN THE COMMITTEE

4.1 Update on Antarctic minke whale body condition

The main results of Konishi and Walløe (In press) had been presented to the Working Group last year (IWC, 2015b). Five response variables related to energy storage and the variable 'sieved stomach content weight' had been systematically analysed using generalised linear models. For all six variables, the analyses had shown a substantial and statistically significant decline over the JARPA period, but no subsequent further decline during the JARPA II period. Last year the Committee had concluded: 'In discussion of these further analyses, the Committee agrees that the analyses which it had requested last year, and those requested by the Review Panel, had been satisfactorily completed.' At this year's meeting, Konishi and Walløe (In press) presented these analyses in more detail and with a number of diagnostic plots, together with results which were similar to those obtained during the meeting last year. These results indicated that important changes took place in the Antarctic ecosystem during the 1990s. The authors argued that the most important cause of the changes was most likely to have been the simultaneous increase in numbers of other krill feeders, especially humpback whales.

In discussion, clarification was sought regarding the form of the relationship to day of the year in the models, and how this had been selected. Analyses presented previously had suggested that BIC was lower when the model assumed a linear relationship rather than the quadratic relationship presented in Konishi and Walløe (In press). Walløe reported that various forms for this relationship in the model had been investigated, ranging from linear to cubic, and that the quadratic form had the lowest BIC.

De la Mare and McKinlay held the view that the real issue was the heterogeneous manner in which the data were collected, and disagreed with the statement from last year's meeting that the analyses requested by the Working Group and later by the Expert Panel had been satisfactorily completed. In particular, they considered that the following points had not been fully addressed:

(1) develop a conceptual model of the system under consideration;
(2) use the conceptual model to identify a set of covariates to consider in the modelling;
(3) start with a 'full model' and base selection of which factors to include and of which of their interactions to treat as random effects on a reduction process; and

(4) apply both Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) as model selection criteria to simplify models and examine the sensitivity of results to the different models selected.

SC/66a/EM01 discussed the relative merits of the two most widely used information criteria, AIC and BIC, in the process of statistical model selection. At recent Scientific Committee meetings the question of whether to prefer AIC or BIC had been a matter of contention in the analysis of Antarctic minke whale nutritive condition. The authors, drawing on both the statistical literature and results from a simulation experiment conditioned on analyses of minke whale nutritive condition, provided some recommendations on appropriate practice in the development and presentation of statistical analyses that use model selection. They concluded that the choice of which information criterion to use depends on the purpose of the analysis, the sample size, and the specifics of the realised experimental design. There is no simple answer that BIC is always better or worse than AIC. They stressed that always preferring BIC over AIC by relying on the asymptotic consistency property of BIC to select an appropriate model as sample sizes become very large is reasonable only if the sample sizes are very large in the context of the model being developed. In the specific case of analyses of minke whale nutritive condition, simulation results conditioned on the realised spatial and temporal sampling patterns of data collected during JARPA indicated that, based on the actual sample sizes, it was likely that models selected using BIC under-estimated the complexity necessary to adequately capture the main features of the data.

The members of the Working Group thanked the authors of SC/66a/EM01 for the practical advice on appropriate practice for model selection, which should be broadly useful in statistical applications.

In response, Konishi and Walløe argued, with reference to statisticians including Adrian Raftery and Nils Lid Hjort, that BIC should be preferred over AIC to be able to select the best model out of a number of possible models in a complex situation with many potential explanatory variables, interaction terms and random effect terms, and a large number of data points. AIC has a positive probability of overestimating the true dimension, even asymptotically, and in practical statistical work AIC tends to overestimate the number of parameters needed. As an illustration, Konishi and Walløe repeated the model selection procedure used in Konishi and Walløe (In press) for the dependent variable Fat Weight, but using AIC as a selection criterion instead of BIC, which was used in the paper. Three new terms were included in the final AIC model, two random effect terms and one ordinary categorical variable, and the degrees of freedom increased from seven to 21. However, for the AIC-based model the fat weight declined over the JARPA years by 9.1kg/year (SE=2.6), which was not very different from the 8.3kg/year (SE=1.4) obtained from the much simpler model using BIC.

In discussion Zeh noted that the goals for any modelling exercise need to be clear when considering a model selection metric, as models selected using the more conservative BIC statistic would provide more robust parameter estimates for a reduced model compared to models selected with the more permissive AIC statistic. In addition to these considerations, it was noted that the fact that these approaches to model selection provide similar results was encouraging.

De la Mare and McKinlay (see Appendix 2) outlined past recommendations of the Scientific Committee and the Working Group prior to the meeting. The authors considered that points (1)-(4) above had not been fully addressed. A conceptual model of the accumulation of body condition in Antarctic minke whales during the feeding season was set out in brief, and an explanation given for why interaction terms were needed in models to account for important features of the data. Some illustrative figures from de la Mare *et al.* (2014a), presented to the JARPA II Review (IWC, 2015a), were extracted to outline the basic properties of the data, the nature of the statistical correction to the data, and that models selected by AIC included interaction terms that allowed important features of the data to be accounted for. Of particular importance was that different sectors of the Antarctic have different apparent trends in body condition. As a contrast, the properties of models (favoured by BIC) that did not include interaction terms were also illustrated; these produced the same year trend in every sector but with different offsets. The authors considered that this class of model did not reflect important properties of the data.

De la Mare, McKinlay and Double, as those authors present of de la Mare *et al.* (2014a; 2014b) submitted to the JARPA II Review, re-iterated the conclusions of those papers that the analyses of minke whale body condition based on JARPA/JARPA II data were insufficient for determining whether there had been trends in Antarctic minke whale body condition or otherwise. The biology of the Antarctic minke whale and the highly variable Antarctic environment lead to within season spatial and temporal heterogeneity in the distribution of individuals by sex, age, maturity, and condition. Consequently, attempting to monitor subtle trends in biological parameters of Antarctic minke whales such as blubber thickness and stomach contents requires a carefully considered and consistent sampling strategy. They had shown that JARPA and JARPA II had not followed such a strategy and consequently the data are confounded and lead to biased estimates. It is difficult to separate the effects of any possible real trends with the confounding effects of the systematic change in the spatial and temporal patterns of data collection over the years of JARPA I/II. By way of example, de la Mare (2012) had demonstrated that analyses using data collected during JARPA of parameters that should be close to constant, such as sex ratio, had both biases and spurious trends. Even after complex statistical modelling, any long-term trends inferred from these data are likely to be imprecise and unreliable. Standard linear and linear mixed-effects models had been applied to attempt to correct for heterogeneity in the manner in which the data were collected, and these had confirmed that interaction terms were significant. While some models had suggested statistically significant trends in some body condition parameters others had not. The results of these analyses indicated that spatial and temporal heterogeneity requires complicated statistical modelling and that statistical inferences from these models are weak and ambiguous.

In response, Konishi and Walløe stated that McKinley and de la Mare had presented their complex 'full model' (d3) in SC/66a/EM01. Konishi and Walløe considered this model to be far too complex to be a reasonable 'full model', but even so they had tried to reduce this model with FatWeight as the dependent variable as instructed by Zuur *et al.* (2009) and the *ad hoc* group appointed by the EM Working Group at the previous year's meeting, using both BIC and AIC. However, in this complex model it was not possible first to add the five random effects one at a time. With BIC the resulting model selected was:

FatWeight ~ DateNum:Sex + YearNum + BLm + DateNum

The coefficient for YearNum was -8.08kg/year (SE=1.49) with a t-value of -5.43, and df=7. This slope was similar to what Konishi and Walløe had found previously. When AIC was used, the resulting best model selected was:

FatWeight ~ DateNum + DateNum:Year:LongNum + YearNum:LonSect + BLm + Sex + Year

There was no explicit slope in the absence of the YearNum variable.

However, since this complicated 'full model' did not contain YearNum as an explicit possibility, Konishi and Walløe added YearNum to de la Mare and McKinlay's d3 model. The starting model was then:

FatWeight ~ DateNum*Sex + DateNum*Year:LonSect:LongNum + DateNum:LonSect:Ice + YearNum:LonSect + BLm +YearNum

Reduction using BIC gave the same reduced model as above. Reduction by AIC gave:

FatWeight ~ DateNum + Sex + Year:LonSect:LongNum + DateNum:LonSect:Ice + BLm + YearNum + Year

with coefficient for YearNum = -1.48kg/year (SE=0.59), p=0.01294, df=77, but 66 coefficients were not defined because of singularities. Konishi and Walløe therefore concluded that both BIC and AIC showed a statistically significant decline in fat weight (by 9% and 2%, respectively, over the JARPA years), but that in addition there was geographical and temporal heterogeneity.

General discussion included de la Mare stating that he did not agree that the data supported the conclusion of a change in a trend (if present) between the JARPA I and JARPA II periods. Zeh considered that the BIC selection criterion was more appropriate than AIC for the models under consideration. Walløe held the view that all of points (1) to (4) above listed by de la Mare and McKinlay had been addressed. Butterworth stated that differences in view seemed to hinge on whether or not the choice of the 'full model' on which the last year's meeting had based its conclusions was appropriate, but drew attention to the fact that the *ad hoc* group had considered this had included all terms which they had deemed biologically realistic *a priori*, and that this group had included a member of the review panel.

Butterworth asked whether demonstrating that the model selected evidenced a trend with time statistically significant at the 5% level was the most pertinent question. As far as he recalled, for the results which had been shown over time across various model and selection criterion choices, a robust feature was that point estimates of trends with time over the JARPA period were consistently negative. Differences were related to standard error estimates which were typically larger when more complex models were selected, so that on occasions the 5% significance level was not attained. However in using the results from these analyses for input to ecosystem models, it would be the trend and associated standard error estimates that would be needed, and these could defensibly be used even if the trend estimate itself failed to attain significance at the 5% level.

De la Mare agreed in certain respects, stating that in the application of the results of analyses of body condition in the development of ecological models several issues need to be addressed, particularly relating to spatial scales and differing apparent trends in nutritive condition in different regions and across a range of species, including how to infer possible trends in important species for which there are very few data. For this purpose, model analyses should be directed towards capturing the important features of the data, in which case statistical inference is a secondary issue. When the objective is to make accurate predictions he considered that more complex models with interaction terms need to be included.

Walløe nevertheless considered that the matter of the presence or otherwise of a declining trend remained an important issue in understanding the behaviour of the Antarctic ecosystem. De la Mare explained his view was that while the possibility of such a decline was not excluded, the analyses by himself and his colleagues had indicated that the data were also open to different interpretations.

There was not sufficient support in the Working Group to modify its conclusion (subsequently endorsed by the Committee) from last year that 'a decline in blubber thickness and in fat weight that was statistically significant at the 5% level had occurred'. De la Mare and McKinlay considered that last year's conclusion was premature because in their view it was not based on the full analyses recommended by the JARPA II Expert Review Panel. The Working Group expressed nevertheless appreciation to all those who provided analyses to the meeting for their substantial contributions.

Given earlier recommendations by the Committee and the continuing debate of how best to model the data, the Working Group **recommended** that additional analyses be undertaken on both the blubber thickness and body fat data. It **encouraged** the various scientists involved in these analyses to collaborate to develop a set of models that best capture the Committee's previous recommendations, taking into account the structure of the underlying processes giving rise to the data. To facilitate this, the Working Group suggested that the interested scientists apply for access to the data under Procedure B of the Data Availability Agreement; it **requested** the data holders to consider such requests favourably.

4.2 Case studies of the effects of long-term environmental variability on whale populations

4.2.1 Compilation of long time series of cetacean demographic parameters and/or abundance and potentially relevant environmental variables

Last year an intersesssional Correspondence Group had been appointed to continue with the compilation of cetacean-related and environmental time series to address the effects of long-term environmental change on cetaceans. The Correspondence Group reported progress on modelling of two populations with respect to detection of environmental effects: southwest Atlantic right whales and North Pacific gray whales (see Annex F, items 3.1 and 4.1) but the time has been taken up refining the demographic models so as to clearly identify which biological parameters and which points in the life cycle are affected by environmental variability. In the case of southwest Atlantic right whales, calf survival rate was found to be the parameter most affected, when cryptic mortality (calf dies without being seen) is included. Potentially relevant environmental indices have been compiled, and the next step is to investigate correlations. Data sets on other baleen whale species, as identified in IWC (2015b), will also be examined for their potential to reveal effects of environmental variability.

4.2.2 Review the Report of the IWC Climate Change Steering Group Meeting

Last year the Scientific Committee had established a Steering Group to coordinate and organise a meeting of the Steering Committee on Climate Change, which took place on 19 August 2014 at the University of Glasgow. The purpose of this meeting was to review work by the Committee to date, place this in the context of latest knowledge, and produce a plan for an ongoing work programme by the Committee. The report of this meeting (SC/66a/Rep07) was received and reviewed during a joint E/EM session, and the discussions are fully presented under Annex K, Item 10.1. Of note to this Working Group, among the recommendations from this report is 'forming a cross-cutting theme with various of the Scientific Committee's sub-committees including, but not necessarily limited to, Ecosystem Modelling (EM)'. The E Sub-committee **agreed** to maintain this Steering Group as an intersessional Correspondence Group to continue their work to identify climate change priorities within the Committee (see Annex K, Table 2).

4.3 Competition among baleen whales: how can we measure and model it?

This subject had been an area of emphasis at last year's meeting, both from a modelling and empirical perspective. The Committee had agreed that there was a critical need for species-specific, fine-scale data on cetacean feeding and prey to provide parameters for individual-based models of competition between baleen whales. The Committee had also noted an urgent need to develop the analytical and modelling tools to scale from individual-based whale foraging scales to broad spatial scales across species and ecosystems, using information about baleen whale energetics and feeding functional forms, as well as existing satellite tag, spatial and temporal data. Finally, the Committee had also recognised that the development of competition models should be conducted in parallel with data collection because the models can inform data collection and experimental design, and *vice versa*.

This year the Working Group received a progress report on an IWC-supported project entitled 'using baleen whale tag data to inform ecosystem models of competition' (SC/66a/EM04). The project addressed the need identified during SC/65b to model the potential for competition among baleen whales, and to use existing data on cetacean feeding to provide parameters for individual-based models. During the first year of the project, Friedlaender and co-workers have developed tools and results that can be used at a range of spatial scales and across species. Using methods from Goldbogen *et al.* (2011), they presented new information on the energetic costs of feeding for all rorqual whale species and how these change for each species as a function of prey density. This effort has yielded significant insights that will be used to parameterise individual-based models in year two of the study. They also presented results from a state-space switching model (SSM) applied to satellite-tag data from humpback and minke whales in the Antarctic. The goals of this analysis were to provide estimates for the proportion of time spent in different behavioural states (area-restricted search *vs* transit) for the two species across broad spatio-temporal scales. This analysis will continue with a more robust habitat modeling exercise to compare the environmental parameters that define habitat where area-restricted search behaviour is observed for each species and eventually to generate ecological niche models and amount of overlap between the two species. This information will be used in concert with individual-based models to generate estimates for the potential of competition between sympatric baleen whale species (see Item 3 above).

In discussion of the broad-scale aspects of the project it was noted that SSMs applied to Argos tracks to infer movement behaviour do not always capture true behaviour since transit does not preclude foraging even if this is not revealed by the SSM. This appears to be the case in minke whales, who have lower energetic demands associated with feeding than other larger baleen species and can 'eat on the go', whilst transiting and 'hiding out' are likely important predator (killer whale) avoidance strategies. Furthermore, minke whale migration behaviour may reduce their energy requirements when compared with the long-range migrations undertaken by other baleen whales. The Working Group concluded that SSMs provide a useful framework for ecological analyses, but that further work is needed to improve their application through the use of satellite tags that provide higher resolution information about movement and foraging behaviour.

Mention was made that stomach contents from lethal sampling would complement these studies by providing information on the actual prey items consumed. However, the Working Group noted that active acoustic sampling is conducted alongside net tows to map prey in the vicinity of tagged individuals and, further, that individual-based tags are now being employed which integrate video cameras to allow direct observation of the prey being consumed, alleviating some of the ambiguity surrounding the acoustic methods.

The direct measurement of feeding rates and the derivation of functional relationships between foraging effort and prey concentration using energetic models opens new possibilities for generating ecosystem-level information. The question arose of whether it is currently possible to estimate consumption rates and efficiency in an area, along with the uncertainty surrounding this estimation? The Working Group was informed that it is currently very difficult to derive consumption rate estimates because echosounder measurements of prey fields provide information that allows estimation of prey density and biomass but not abundance. Therefore, it is not yet possible to resolve the impact of an individual whale on prey fields.

A related question was whether these functional relationships derived from biomechanistic models can be used to 'invert' the problem and infer krill density from the observed foraging effort (lunges per day) in cases where prey measurements are not available? In response it was noted that this does indeed seems possible, given that the models provide information about the threshold prey concentrations required for optimal foraging in relation to energy gain and oxygen consumption (Hazen *et al.*, In review). There is sufficient information to do this for humpback whales and fin whales, but minke whales remain data deficient. In this context, de la Mare reported that the next development of the IBEM will be to model minke whales (see Item 3 above).

Finally, the question was raised as to whether sex-/age-specific information could be derived from the observational data? It was noted that pregnant humpback whale females exhibit distinct foraging strategies (hyperphagia) and movement patterns related to the increased energy demands. Being capital breeders, these increased energy demands are primarily associated with preparation for lactation, rather than the pregnancy itself.

4.4 Applications of species distribution models (SDMs)

4.4.1 Preliminary review of SDMs applied to baleen whales

Last year the Working Group had agreed to review the application of species distribution modelling and associated techniques as they pertain to the goals of the Scientific Committee, and established an intersessional Correspondence Group to develop guidelines and recommendations for best modelling practices. This year the Working Group received a progress report on this topic - SC/66a/EM03 provided a preliminary review of SDMs as applied to baleen whales. The general aspects of SDMs were reviewed first, and then SDMs applied to baleen whales were reviewed. SDMs can be categorised into four approaches based on the response variable: presence/absence, presence only, presence/background, and presence/pseudo-absence. Abundance can be used in some models in substitution for presence. SDMs can also be categorized into four methods based on the type of statistical modelling framework: regression, profile, machine learning, and other methods. Several issues such as spatial autocorrelation and collinearity need to be considered prior to modelling. In the second part of the paper, a total of 36 papers published in scientific journals from December 1997 to March 2015 were reviewed. SDMs were applied to all baleen whale species, except pygmy right and Omura's whales. A total of 10 types of statistical models were used in these studies. Although the results significantly contribute to expanding our knowledge of baleen whale ecology, the review concluded that detailed descriptions of the construction and evaluation methods were needed for further consideration of the results. The authors made two recommendations for future studies: that comparison among different SDMs and ensemble modelling should be pursued in future studies; and that appropriate guidelines for parameter settings and evaluation methods for SDMs should be prepared as they are now commonly applied to baleen whales.

The Working Group commended the authors of SC/66a/EM03 for undertaking this review. In discussion, it was **recommended** that the review be expanded to consider guidelines for model diagnostics and examination of residuals. The Working Group observed that few studies describe application of multiple models to the same data for sensitivity analysis. This point was extended to include the idea of ground-truthing and model validation. Simulation was suggested as a cheaper alternative to empirical model validation. An intersessional Correspondence Group was formed to extend the work undertaken by SC/66a/EM03 to include machine-learning techniques (see Item 5, work plan). The Working Group looks forward to receiving a presentation of this review at SC/66b.

4.4.2 Review the Report of the Joint NMFS-IWC Preparatory Workshop 'Towards Ensemble Averaging of Cetacean Distribution Models

A joint IWC-National Marine Fisheries (NMFS) preparatory Workshop titled 'Towards Ensemble Averaging of Cetacean Distribution Models' was held in San Diego, USA, 21 May 2015, in the form of a pre-meeting to SC/66a (see SC/66a/Rep10). As background, in September 2014, the USA's NMFS had hosted a scientific and technical workshop to provide a forum for in-depth discussion on recent work on large whale distribution and occurrence, particularly in waters off the USA's west coast. At the conclusion of that workshop, participants had noted that a number of independent species distribution models (SDMs) had been developed using various methods and data sets, particularly for blue whales and it was agreed that a collaborative effort to develop formal methods to compare and combine predictions from these models was needed. Contemporaneously, a specific interest in developing guidelines and recommendations for best practices in SDMs for large whales had emerged during SC/65b (IWC, 2015b). A Workshop was organised to facilitate that collaborative effort.

The objective of the resultant joint IWC-NMFS Workshop was to convene a group of experts in modelling, statistics, and marine ecology to identify methods to compare and combine model predictions, using existing SDMs for the eastern North Pacific blue whale as a case study. It was anticipated that an ensemble-averaged SDM would have important management applications by showing where blue whales may be more vulnerable to different human activities. This process would lay the groundwork for the future development of models for other large whales (a primary area of interest for NMFS off the USA's west coast, but also more broadly for whale populations of interest to the IWC).

Researchers with relevant models were invited to present on the pertinent aspects of their approaches at the Workshop. Summaries of their presentations are provided in item 2 of SC/66a/Rep10, including a compilation of the characteristics of the models and of the data sets. The Workshop **agreed** that there was probably a reasonably mature body of literature on model averaging and similar approaches from other fields and **recommended** that a review of such literature should be undertaken with the objective of assisting discussions of the appropriate approach for use within the present blue whale case study. The Workshop noted that, similar to multi-model inference with AIC weights, the composition of the candidate set of models can be influential on the resulting ensemble and the outputs which it provides, and determined that the candidate models should be chosen carefully and with transparency about the degree of similarity between them. The Workshop **agreed** that the development of a meta-data collection for each candidate model for an ensemble is necessary. The metadata would contain information on key management questions; spatial and temporal scales; how error was estimated and propagated, and whether correlation structure of errors had been taken into account for details about source datasets, modelling assumptions, etc. The Workshop recognised that there had been insufficient time to consider the issue of model validation or testing although some options were discussed briefly. It **agreed** that further review and consultation on methods for model validation should be undertaken as part of the preparatory requirements to conduct an ensemble averaging exercise at a future workshop.

The Workshop considered a number of different methods for ensemble modelling, including the 'bounding box' approach (i.e. treating ranges of predictions from models as realisations of a probability distribution) and the Bayesian framework, using the existing predictive output, but further discussion on those and other methods is required. Regardless of the method ultimately used, combinations and comparisons of models must be made on identical units (i.e. whether relative or absolute densities or probabilities of occurrence) and/or management implications.

In recognition of the need to develop methods to average different model types, the Working Group **recommended** a review of scientific fields such as climate change research for methods to combine disparate model types. The Working Group thanked the participants and organisers of the Workshop for this valuable contribution to species distribution modelling for cetaceans.

The Steering Group for this Workshop was re-appointed to continue to advance the agenda and set out objectives for a following Workshop (see Item 5, work plan).

4.5 Other

During SC/65b the Working Group had recommended that attention should be paid to synoptic surveys/studies of krill density and its effect on multi-species competition (including the impact of the krill fishery). This year the Working Group received SC/66a/EM05 (given via video-conference) describing preliminary analyses to characterise the foraging grounds of Antarctic blue whale (*Balaenoptera musculus intermedia*) during a recent joint New Zealand-Australia Antarctic Ecosystems Voyage (see SC/66a/SH07 for further details). The development made possible by the combination of active (echo sounders) and passive acoustics (sonobuoys) was the ability to find aggregations of blue whales using passive acoustics and then use the active acoustics to measure the characteristics of krill swarms, both within the blue whale aggregations and in the surroundings. One of the striking results was that although krill swarms were widely distributed, the blue whales were highly aggregated in a relatively small area. Preliminary results from this survey also suggested that krill swarms were smaller and denser inside vocal aggregations of Antarctic blue whales, but had a similar encounter rate outside vocal aggregations. Bearing in mind that calling blue whales can be detected over hundreds of kilometres, only one aggregation of roughly 80 whales was detected within the large area surveyed, and that aggregation appeared to persist for the several weeks that the ship was in the region. This study demonstrated that using these two complementary technologies provides insights into sub meso-scale Antarctic blue whale foraging behaviour. Further work is planned to analyse the data from additional surveys and to model Antarctic blue whale densities from passive acoustics.

The Working Group welcomed this presentation, and congratulated the authors for their informative results. In discussion, it was noted that this study demonstrated the feasibility of combining passive acoustics to locate blue whales, and active acoustics to study their prey field. It was also noted that by combining passive and active acoustics, studying the foraging ecology of blue whales in the Southern Ocean is now a much more tractable proposition. These methods will assist in characterising medium-scale foraging ecology, where the fine-scale is covered by suction-cup multi-sensor tagging, and the broad-scale is covered by synoptic prey surveys and whale satellite tracking. The Working Group **encouraged** further analyses of the data from this study.

5. WORK PLAN AND BUDGET REQUESTS

A Steering Group was formed to plan next year's Working Group agenda and to identify potential Invited Participants intersessionally (Table 1). The Working Group **agreed** that its work plan before the 2016 Annual Meeting would be as follows.

(1) **Joint CCAMLR-IWC Workshop.** A Steering Group with members from both IWC/SC and CCAMLR was formed to develop terms of reference for the joint Workshop intersessionally and consider participation and contributions to the Workshop against these (see Table 2). The Steering Group also should consider the ongoing efforts by this Working Group in the compilation of long time series of cetacean and environmental variables, as these data may be relevant to the objectives of the Workshop.

(2) **Antarctic minke whale blubber thickness and body fat analyses.** The Working Group **encouraged** collaboration among the scientists involved in these analyses to develop a set of models that best capture the Committee's previous recommendations. It **suggested** that the interested scientists apply for access to the data under Procedure B of the Data Availability Agreement, and **requested** the data holders to consider such requests favourably.

(3) **Climate change and compilation of long time series of cetacean and environmental variables.** The Sub-committee on Environmental Concerns **agreed** to maintain the Steering Group on climate change as an intersessional Correspondence Group (see Annex K, table 2) to form a crosscutting theme with several of the Scientific Committee's sub-committees including this Working Group. In addition to possible applications in ecosystem models during the joint IWC-CCAMLR Workshop, the ongoing compilation of long time series of cetacean and environmental variables by the Working Group is also relevant for climate change purposes.

(4) **Competition among baleen whales.** Regarding the topic of measuring and modelling competition among baleen whales, the Working Group **encouraged** the completion of a manuscript relating to the empirical identification of different patterns in baleen whale foraging and movement, and the formulation of individual-based energetics models for humpback and mike whales using information generated from the new observational data sets.

(5) **Applications of species distribution models (SDMs):** (a) an intersessional Correspondence Group (see Table 4) will continue the review of applications of species distribution models in the context of requirements within the Committee in order to develop guidelines and recommendations for best modelling practices. This Correspondence Group will also review the literature on machine learning methods (boosted regression trees, random forests, MaxEnt, support vector machines, etc.) and develop a guideline, to be completed by October 2015. (b) The Steering Group appointed for the planning the joint NMFS-IWC Preparatory Workshop 'Towards Ensemble Averaging of Cetacean Distribution Models' (see Table 5) was re-appointed to continue advancing the objectives set out for the following workshop. To facilitate progress, four sequential phases were envisioned as follows.

Phase 1

Form three intersession Correspondence Groups (see Table 5; membership to be confirmed) to conduct the following tasks:

CORRESPONDENCE GROUP 1

Review statistical literature and report on techniques that are available and appropriate for building ensemble models.

CORRESPONDENCE GROUP 2

Conduct comparable sensitivity analyses to understand the effect of prediction scale for models to be considered for the ensemble.

CORRESPONDENCE GROUP 3

Identify potential management objectives/advice required (including whether there is a need for real-time predictions *vs* long-term averages).

Phase 2

Members of the Correspondence Groups to complete their projects by February 2016 when the products will be presented to all Correspondence Groups and other invited participants. At this time, plans will be developed for finalizing the products and individuals will also be identified to create a draft plan for developing the ensemble model.

Table 1
Intersessional Steering Group for the preparation of next year's Working Group agenda.

Group	Terms of Reference	Membership
EM planning (Steering Group)	Solicit contributions, liaise with prospective Invited Participants.	Butterworth, de la Mare, Friedlaender, Kitakado (Convenor), Punt, Walløe

Table 2
Steering Group for the preparations for development of a relationship between IWC and CCAMLR.

Group	Terms of Reference	Membership
Joint IWC SC/EM and SC-CCAMLR WG-EMM facilitation and planning (Correspondence Group)	Continue formalisation of a relationship between the IWC and CCAMLR; commence planning of a joint Workshop between IWC SC/EM and SC-CAMLR WG-EMM.	Butterworth (IP), Currey (New Zealand/SC-CCAMLR/IWC/SC Observer), de la Mare (Australia), Ichii (Japan), Kawaguchi (CCAMLR Convenor), Kitakado (IWC Convenor), Kock (Germany), Kovacs (Norway), Trathan (UK), Watters (US/SC-CCAMLR WG-EMM/IWC/SC EM Observer), Thomas

Table 3
Intersessional Correspondence Group on the effects of long-term environmental variability on whale populations.

Group	Terms of Reference	Membership
Effects of long-term environmental variability on whale populations (Correspondence Group)	Identify long time series (≥20 years) of cetacean demographic parameters and/or abundance; identify potentially relevant environmental time series.	Cooke (Convenor), de la Mare, Palacios

Table 4
Intersessional Correspondence Group for the discussion of applications of species distribution models in the context of IWC/SC requirements.

Group	Terms of Reference	Membership
Applications of species distribution models (SDMs) (Correspondence Group)	Develop guidelines and recommendations for best practice in modelling steps.	Friedlaender, Kelly, Kitakado, McKinlay, Murase (Convenor), Palacios, Palka

Table 5
Steering Group Towards Ensemble Averaging of Cetacean Distribution Models and associated Correspondence Groups.

Group	Terms of Reference	Membership
Towards Ensemble Averaging of Cetacean Distribution Models (Steering Group)	Advance objectives of Ensemble Averaging Distribution Model Workshop	Becker, DeAngelis, Kitakado, Palacios, Redfern
Correspondence Group 1	Review statistical literature and techniques	Palacios (Convenor)
Correspondence Group 2	Conduct sensitivity analysis	Becker (Convenor)
Correspondence Group 3	Identify potential management objectives/advice required	DeAngelis (Convenor)

Phase 3
Members of the Correspondence Groups to have final products and a draft plan for developing an ensemble model complete by May 2016 when a following workshop will be convened to share the products and plan to members. Possibilities of holding a following include either a pre-IWC meeting or a separate workshop to develop plans for developing the ensemble model. [Note: In subsequent discussions during this year's meeting, this Working Group noted that conflicts with other pre-meetings and workshops prior to SC/66b might cause the Ensemble Model Workshop to be postponed until SC/67a (2017)].

Phase 4
The ensemble model will be developed. This model will be presented at SC/67 (2017).

Although, NMFS and IWC are identified as the primary source for this funding, additional sources of funding will need to be identified to conduct the ensemble modelling effort, as part of the strawman plan developed during Phase 2.

(6) **Other.** The Working Group **encouraged** further analyses of the data resulting from the acoustic survey of blue whales and krill during the joint New Zealand-Australia Antarctic Ecosystems Voyage 2015 (SC/66a/EM05) that could be presented to SC-CCAMLR WG-EMM in 2015 and 2016 and SC/66b in 2016.

Two requests for funding were advanced at SC/65b arising from items (2) and (4). The first request (IWC, 2014a, p.75; 2014e) was a proposal to support fundamental research that will yield parameters derived from field observations of foraging behaviour to be used as input in individual-based energetic models in the context of the RMP in the near term. IWC (2014a) agreed to allocate £4,100 GBP to this work in 2015 and £5,600 GBP in 2016. SC/66a/EM04 reflects progress on this budget allocation. The Working Group welcomed progress to date and **recommended** the work continue.

The second request was to fund the attendance of four Invited Participants in the joint IWC-CCAMLR Workshop

(IWC, 2014a, p.75; 2014d). IWC (2014a) agreed to allocate £4,000 GBP to this work in 2015 and £4,000 GBP in 2016. IWC/66/4(2015) reflects progress on this budget allocation. Amendments reflected in the report have no budgetary implications. The Working Group welcomed progress to date, **recommended** the work continue and noted CCAMLR-XXXIII endorsement of this work.

At SC/65b a request relevant to Item 3 was made by the RMP sub-committee for funding to evaluate density dependent parameters for inclusion in RMP testing based on energetics modelling (IWC, 2014a, p.75; 2014c). IWC (2014a) agreed to allocate £6,000 GBP to this work in 2015 and £6,000 GBP in 2016. SC/66a/EM02 reflected progress on this topic and the Working Group **recommended** continuation of this effort.

6. ADOPTION OF REPORT

The report was adopted on 30 May 2015 at 18:40. The Working Group thanked Palacios for chairing and Bell, Butterworth, Elvarsson, and Kelly for their excellent rapporteurial duties. After four years as Convenor, Palacios announced that he would be stepping down at the end of this year's meeting. He expressed his appreciation to the members of the Working Group for their patience and encouragement during this time, and thanked the other Convenors for their camaraderie and the Secretariat's staff for their tireless support. He indicated that the other Convenors had endorsed Kitakado as a candidate to take on the Convenor role and requested confirmation from the Working Group.

REFERENCES

de la Mare, W.K. 2012. Lurking variables and the interpretation of statistical analyses of data collected under JARPA. Paper SC/64/EM3 presented to the IWC Scientific Committee, June 2012, Panama City (unpublished). 65pp. [Paper available from the Office of this Journal].

de la Mare, W., Candy, S., McKinlay, J., Wotherspoon, S. and Double, M. 2014a. JARPA II sample size calculations ignore spatial and temporal variability. Paper SC/F14/O07 presented to the JARPA II Special Permit Expert Panel Review Workshop, February 24-28 2014, Tokyo, Japan (unpublished). 3pp. [Paper available from the Office of this Journal].

de la Mare, W., Candy, S., McKinlay, J., Wotherspoon, S. and Double, M. 2014b. What can be concluded from the statistical analyses of JARPA/JARPA II body condition data? Paper SC/F14/O06 presented to the JARPA II Special Permit Expert Panel Review Workshop, February 24-28 2014, Tokyo, Japan (unpublished). 58pp. [Paper available from the Office of this Journal].

Eisert, R., Ensor, P. and Currey, R. 2014a. Killer whale studies, McMurdo Sound, Ross Sea, Antarctica, Jan-Feb 2014. Paper SC/65b/SM06 presented to the IWC Scientific Committee, May 2014, Bled, Slovenia (unpublished). 15pp. [Paper available from the Office of this Journal].

Eisert, R., Pinkerton, M.H., Torres, L., Currey, R., Ensor, P.H., Ovsyanikova, E.N., Visser, I.N. and Oftedal, O.T. 2014b. Update on the Top Predator Alliance Proect, 2013-14 season: Killer whales. Report No. WG-EMM-14/52 to CCAMLR. [Available at: *https://www.ccamlr.org/en/wg-emm-14/52*].

Goldbogen, J.A., Calambokidis, J., Oleson, E., Potvin, J., Peyenson, N.D., Schorr, G. and Shadwick, R.E. 2011. Mechanics, hydrodynamics and energetic of blue whale lunge feeding: efficiency dependence on krill density. *J. Exp. Biol.* 214(131-146).

Hazen, E.L., Goldbogen, J.A. and Friedlaender, A.S. In review. Optimal foraging strategies for gigantism: balancing oxygen use and energy gain. *Curr. Biol.*

International Whaling Commission. 2014a. Report of the Scientific Committee. *J. Cetacean Res. Manage. (Suppl.)* 15:1-75.

International Whaling Commission. 2014b. Report of the Scientific Committee. Annex K1. Report of the Working Group on Ecosystem Modelling. *J. Cetacean Res. Manage. (Suppl.)* 15:331-44.

International Whaling Commission. 2014c. Research Proposal for Sub-Committee on the Revised Management Procedure: Evaluation of density dependence parameters for inclusion in RMP testing based on energetics modelling. Lead researchers: de la Mare, W. and Andrews-Goff, V. Paper SC/65b/RMPRP04 presented to the IWC Scientific Committee, May 2014, Bled, Slovenia (unpublished). 1pp. [Paper available from the Office of this Journal].

International Whaling Commission. 2014d. Research Proposal for Working Group on Ecosystem Modelling: Proposal to hold a Joint IWC-CCAMLR Workshop on the Development and Application of Multi-Species Models to the Antarctic Marine Ecosystem. Convenor: Currey, M. Paper SC/65b/EMRP02 presented to the IWC Scientific Committee, May 2014, Bled, Slovenia (unpublished). 3pp. [Paper available from the Office of this Journal].

International Whaling Commission. 2014e. Research Proposal for Working Group on Ecosystem Modelling: Using baleen whale tag data to inform ecosystem models. Lead researchers: Friedlaender, A.S., de la Mare, W. and Goldbogen, J. Paper SC/65b/EMRP01 presented to the IWC Scientific Committee, May 2014, Bled, Slovenia (unpublished). 1pp. [Paper available from the Office of this Journal].

International Whaling Commission. 2015a. Report of the Expert Workshop to Review the Japanese JARPA II Special Permit Research Programme, 24-28 February 2014, Tokyo, Japan. *J. Cetacean Res. Manage. (Suppl.)* 16:369-409.

International Whaling Commission. 2015b. Report of the Scientific Committee. Annex K1. Report of the Working Group on Ecosystem Modelling. *J. Cetacean Res. Manage. (Suppl.)* 16:277-90.

International Whaling Commission and CCAMLR. 2010. Report of the Joint CCAMLR-IWC Workshop to Review Input Data for Antarctic Marine Ecosystem Models,11-15 August 2008, Hobart, Australia. *J. Cetacean Res. Manage. (Suppl.)* 11(2):541-86.

Konishi, K. and Walløe, L. In press. Substantial decline in energy storage and stomach fullness in Antarctic minke whales during the 1990s. *J. Cetacean Res. Manage.*: 34pp.

Krafft, B.A., Krag, L.A., Klevjer, T.A., Skaret, G. and Pedersen, R. 2014. Report from the annual survey of Antarctic krill and apex predator distribution at South Orkney Islands in 2014, and assessing escape mortality of krill in trawls. Report No. WG-EMM-14/16 to CCAMLR. [Available at: *https://www.ccamlr.org/en/wg-emm-14/16*].

Orgeira, J.L., Alderete, M., Jimenez, Y.G. and Gonzalez, J.C. 2014. Long-term study of the at-sea distribution of seabirds and marine mammals in the Scotia Sea, Antartica. Report No. WG-EMM-14/06 Rev.1 to CCAMLR. [Available at: *https://www.ccamlr.org/en/g-emm-14/06Rev*].

Reilly, S., Hedley, S., Borberg, J., Hewitt, R., Thiele, D., Watkins, J. and Naganobu, M. 2004. Biomass and energy transfer to baleen whales in the South Atlantic sector of the Antarctic. *Deep-Sea Research II* 51: 1397-409.

Watkins, J.L., Hewitt, R., Naganobu, M. and Sushin, V. 2004. The CCAMLR 2000 Survey: a multinational, multi-ship biological oceanography survey of the Atlantic sector of the Southern Ocean. *Deep Sea Research* II, 51: 1205-13.

Zuur, A., F., Ieno, E.N., Walker, N.J., Saveliev, A.A. and Smith, G.M. 2009. *Mixed Effects Models and Extensions in Ecology with R*. Springer. 574pp.

Appendix 1

AGENDA

1. Introductory items
   1.1 Convenor's opening remarks
   1.2 Arrangements for the meeting
   1.3 Election of Chair
   1.4 Appointment of rapporteurs
   1.5 Adoption of agenda
   1.6 Documents available
2. Review ecosystem modelling efforts undertaken outside the IWC
   2.1 Update from CCAMLR's Ecosystem Monitoring and Management Programme (WG-EMM) on krill and its dependent predators
   2.2 Update on planning for joint IWC-CCAMLR activities in 2016 and beyond
3. Explore how ecosystem models contribute to developing scenarios for simulation testing of the RMP
4. Review other issues relevant to ecosystem modelling within the Scientific Committee
   4.1 Update on Antarctic minke whale body condition
   4.2 Case studies of the effects of long-term environmental variability on whale populations
      4.2.1 Compilation of long time series of cetacean demographic parameters and/or abundance and potentially relevant environmental variables
      4.2.2 Review the 'Report of the IWC Climate Change Steering Group Meeting'
   4.3 Competition among baleen whales: how can we measure and model it?
   4.4 Applications of species distribution models (SDMs)
      4.4.1 Preliminary review of SDMs applied to baleen whales
      4.4.2 Review the Report of the Joint NMFS-IWC Preparatory Workshop 'Towards Ensemble Averaging of Cetacean Distribution Models
   4.5 Other
5. Work plan and budget requests
6. Adoption of Report

Appendix 2

RECENT RECOMMENDATIONS FOR THE ANALYSIS OF MINKE WHALE BODY CONDITION DATA AND SOME CONSEQUENTIAL PRESENTATIONS TO THE JARPA II REVIEW

W.K. de la Mare and J.P. McKinlay

In its 2014 meeting the SC provided endorsement that analyses of minke whale nutritive condition conducted in the EM Working Group of that year satisfied the requirements for further analysis set out by the Expert Review Panel for NEWREP-A. We have reviewed the text of the Expert Panel and the further analyses that were presented in the EM Working Group, and we unfortunately cannot agree that the Panel's advice has been carried out in full. In particular, we feel the following points remain unaddressed:

(1) develop a conceptual model of the system under consideration;
(2) use the conceptual model to identify a set of covariates to consider in the modelling;
(3) start with a 'full model' and base selection of which factors and their interactions to treat as random effects; and
(4) apply both AIC and BIC to simplify models and examine the sensitivity of results to the different models selected.

Points (1)-(3) are informed by analyses and presentations we have previously provided to the JARPA II and review panel meeting, which we reproduce in part below. We also reproduce the relevant text provided by the Scientific Committee and the JARPA II Expert Review Panel that we feel supports our position. The figures below are drawn from the presentations and papers we provided to the JARPA II Review Workshop.

Report of the Scientific Committee (IWC, 2014, p.45)
Previously, the Committee has requested further analyses of the data, including:

(1) determining whether the models fitted so far capture all the main features of the data;
(2) determining whether the estimate of trend could be made more precise;
(3) analysing the two sexes separately;
(4) including the interaction of slopes by latitudinal band with year as a random effect; and
(5) investigating independence issues by using mixed effects models with trackline as a random effect.

The Committee reiterates its recommendations from previous years that the outstanding issues raised at recent meetings should be examined (for details see main text, Item 4.1). A number of additional suggestions were also made this year. The Committee encourages additional analyses to be undertaken on both the blubber thickness and body fat data and noted that papers should ideally be submitted to the forthcoming JARPA II Review Workshop (see Item 17.3).

Excerpts from: Report of the Expert Workshop to Review the Japanese JARPA II Special Permit Research Programme (IWC, 2015): *Para 8.3.2*
The factors considered in the models on which the analyses in Konishi and Walløe (2014) and Konishi *et al.* (2014)

are based are not derived from biological hypotheses, but rather arise primarily from discussions within the Scientific Committee. The Panel **recognises** that this is a consequence of the nature of the discussions within the Scientific Committee rather than a failure by the authors. The Panel **recommends** that the authors of Konishi and Walløe (2014) and Konishi *et al.* (2014) first develop a conceptual model of the system under consideration and use that to identify a set of covariates to consider in the modelling. Model selection should always be guided by underlying knowledge of the system. It is therefore inappropriate to automatically select the 'best model' because such a model can lead to covariates being selected for which there is no reason that there are related to response variable. Following the selection of which factors to consider in the modelling, the following steps should be undertaken:

(1) identify whether any of the covariates are highly correlated and either: (a) exclude a subset of the covariates so that the remaining covariates are uncorrelated; or (b) develop new covariates which represent independent aspects of the current covariates (using for example PCA);
(2) select a 'full model' (this may be difficult if the data set is unbalanced) and base selection of which factors and their interactions to treat as random effects - the models should be fitted using REML and a model selection approach such AIC, BIC or standard hypothesis testing approach applied;
(3) select the fixed effects structure given the random effects structure selected at step (2), where the models are fitted using maximum likelihood; and
(4) use REML to fit the best model identified in (3) above.

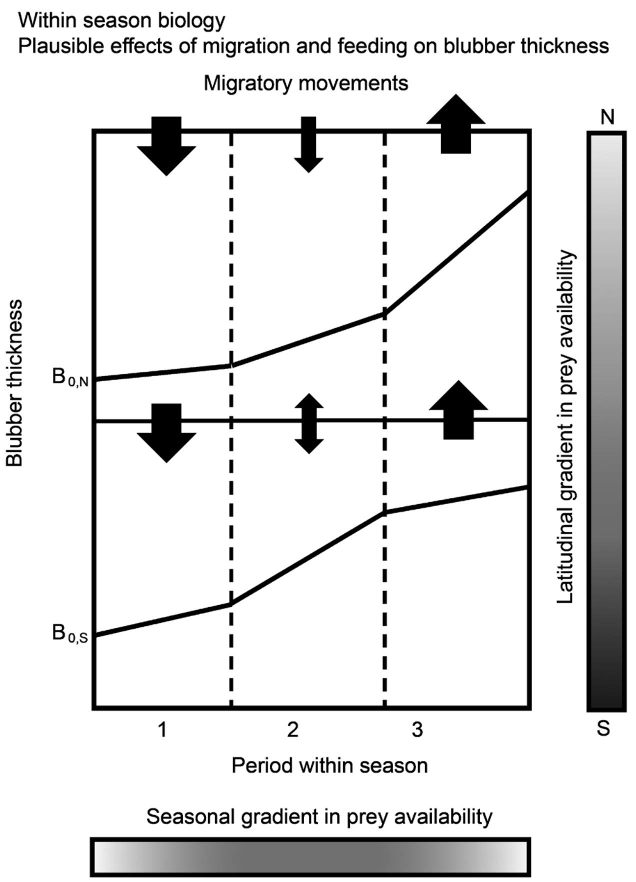

Fig. 1. Conceptual biological model of in-season blubber thickness underlying the de la Mare *et al.* (2014a; 2014b) analyses.

The Panel also recommends that future analyses of the data on the condition of Antarctic minke whales include: (a) consideration of a model in which year is a categorical variable and is treated as a random effect if a plot of residuals against year show there are residual patterns by year; and (b) examination of how robust the results in Konishi *et al.* (2014) are to basing model selection on BIC rather than AIC.

Given the discussion in a number of papers presented by both proponents and observers, the Panel offers the following summary of the merits or otherwise of AIC and BIC, both commonly used metrics for model selection. The two metrics, while mathematically similar (twice the negative log-likelihood plus a penalty term), arise from different underlying arguments (Kass and Raftery, 1995). BIC will tend to select simpler models than AIC if the number of data points exceeds seven, all things being equal (because the penalty term for additional parameters is larger than for AIC). Simulations have shown that AIC often selects a more complex (wrong) model over the simpler (and correct) model (Kass and Raftery, 1995). Whether this necessarily means that BIC is always better than AIC is not clear because AIC attempts to find the best approximating model rather than true model (which would be rarely in the set of candidate models). BIC is, however, generally preferred to AIC for 'large' data sets. However, what constitutes 'large' in any particular case depends inter alia on the set of models under consideration. Thus, for any one problem the selection between AIC and BIC is seldom definitive. Many practicing statisticians consequently often apply both AIC and BIC, examine the sensitivity to the different models selected and apply expert judgement.

8.2 Observer summary of de la Mare *et al.* (2014a; 2014b) and Double *et al.* (2014)

The authors of de la Mare *et al.* (2014a; 2014b) summarised the conclusions of their analyses of minke whale body condition that used the JARPA/JARPA II data. The biology of the Antarctic minke whale and the highly variable Antarctic environment are likely to generate within season spatial and temporal heterogeneity in the distribution of individuals by sex, age, maturity and condition. Consequently attempting to monitor subtle trends in biological parameters of Antarctic minke whales such as blubber thickness and stomach contents requires a carefully considered and consistent sampling strategy. They showed that JARPA and JARPA II have not followed such a strategy and consequently the data are confounded and biased. Even after complex statistical modelling, any long-term trends inferred from these data are likely to be imprecise and unreliable. Standard linear and linear mixed effects models were applied to attempt to correct for heterogeneity in the manner in which the data were collected and these confirmed that interaction terms were significant. The results of these analyses indicated that spatial and temporal heterogeneity requires complicated statistical modelling and that statistical inferences from these models are weak and ambiguous. While some models suggest statistically significant trends in some body condition parameters others do not, with the conclusion that the data collected during JARPA and JARPA II are insufficient for determining whether there have been trends in Antarctic minke whale body condition or otherwise.

The authors pointed out that backwards model selection using BIC in the nutritive condition analyses leads to the illogical consequence that a valid *a priori* hypothesis for spatial and yearly interaction terms is accepted, but would then subsequently be discarded even though it is highly statistically significant. Furthermore the models using interaction terms have the lowest AICs.

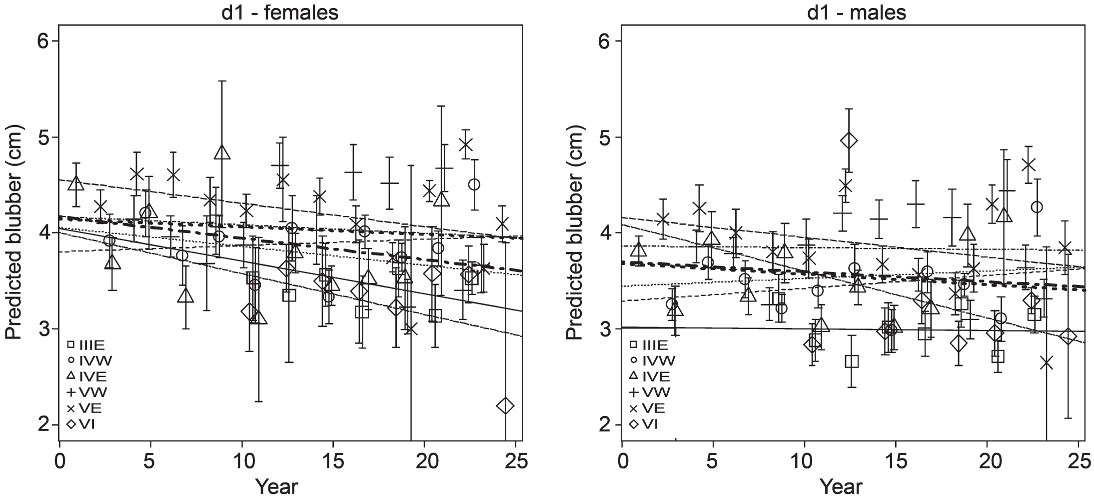

Fig. 2. An overview of the properties of the blubber thickness data showing the mean blubber thickness by year without any other statistical modelling or corrections. The data are stratified by half area. Different strata have different apparent trends – some are apparently statistically significant – some are not.

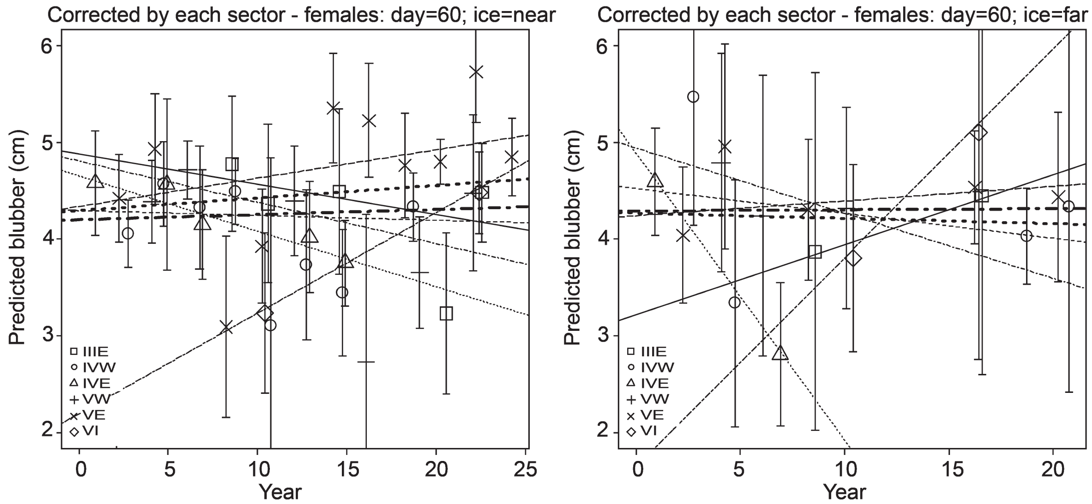

Fig. 3. Not all the data are collected at the same time of year so to compare across years the blubber thickness needs to be corrected to a 'standard date'. These figures illustrate how what that standardisation would lead to if a separate slopes and intercepts model on blubber thickness by date was used to calculate the blubber thickness at mid-season.

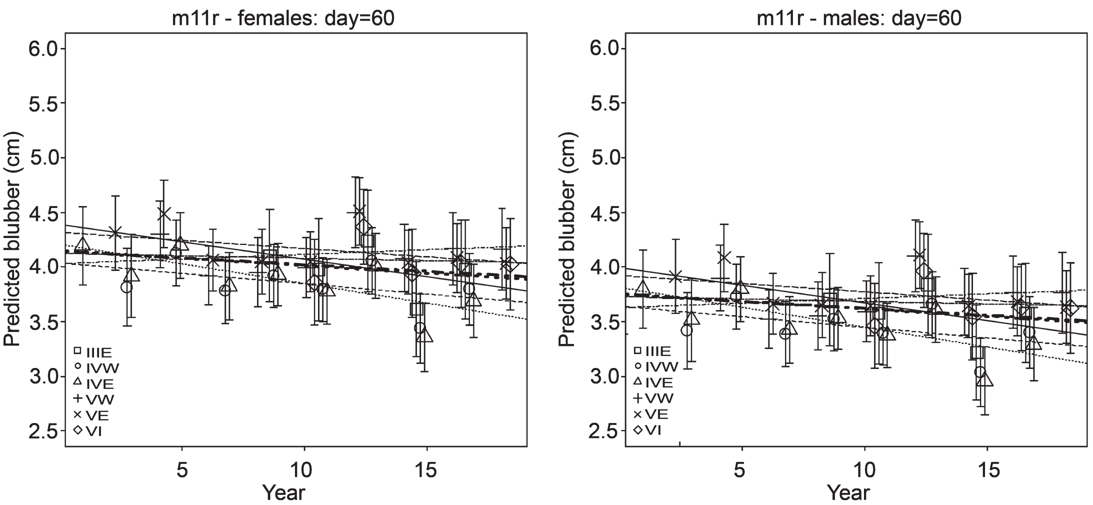

Fig. 4. However, statistical modelling can reduce the variability by taking into account more covariates. These graphs show predicted blubber thicknesses from the best AIC model with random effects and interaction terms. This model captures one of the main features of the data that different strata have different trends.

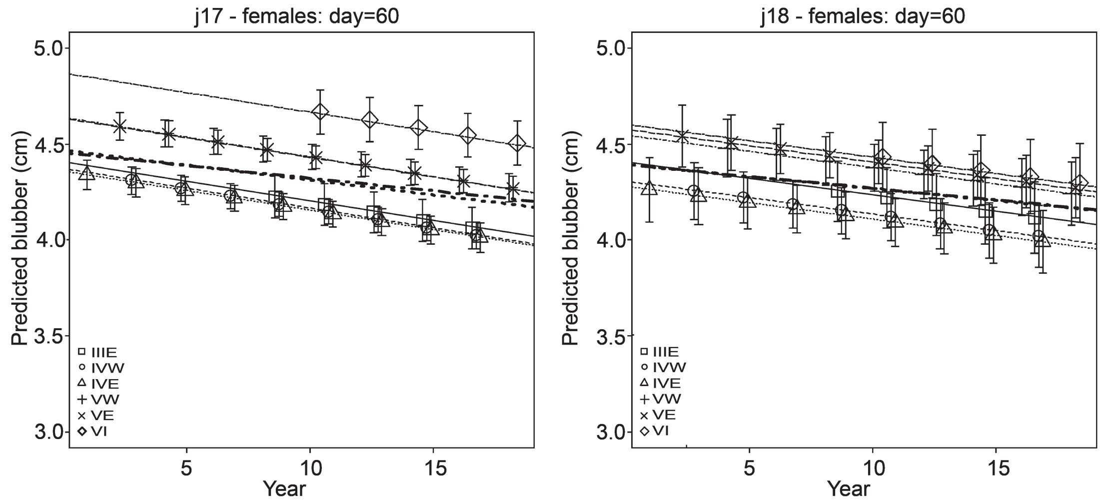

Fig. 5. These are the types of model selected by BIC, where interaction terms are excluded. The model now produces a series of lines with the same slope by year with only but with separate intercepts. This model does not capture the main features of the data.

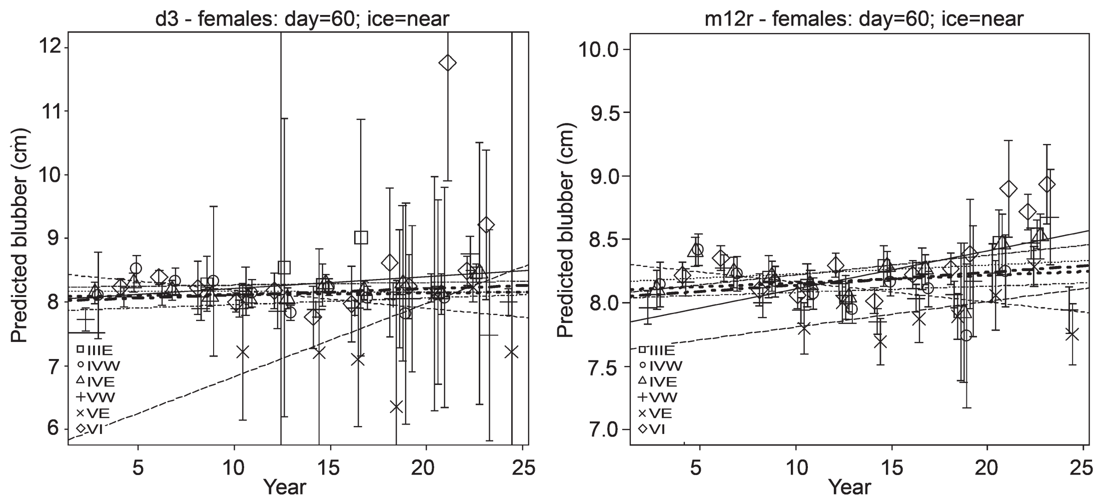

Fig. 6. An additional analysis – trends in total body mass appear to be increasing.

REFERENCES

de la Mare, W., Candy, S., McKinlay, J., Wotherspoon, S. and Double, M. 2014a. JARPA II sample size calculations ignore spatial and temporal variability. Paper SC/F14/O07 presented to the JARPA II Special Permit Expert Panel Review Workshop, February 24-28 2014, Tokyo, Japan (unpublished). 3pp. [Paper available from the Office of this Journal].

de la Mare, W., Candy, S., McKinlay, J., Wotherspoon, S. and Double, M. 2014b. What can be concluded from the statistical analyses of JARPA/ JARPA II body condition data? Paper SC/F14/O06 presented to the JARPA II Special Permit Expert Panel Review Workshop, February 24-28 2014, Tokyo, Japan (unpublished). 58pp. [Paper available from the Office of this Journal].

Double, M., de la Mare, W. and Gales, N. 2014. Observers' Statement to the JARPA II Special Permit Expert Panel Review Workshop. Paper SC/F14/O08 presented to the JARPA II Special Permit Expert Panel Review Workshop, February 24-28 2014, Tokyo, Japan (unpublished). 4pp. [Paper available from the Office of this Journal].

International Whaling Commission. 2014. Report of the Scientific Committee. *J. Cetacean Res. Manage. (Suppl.)* 15:1-75.

International Whaling Commission. 2015. Report of the Expert Workshop to Review the Japanese JARPA II Special Permit Research Programme, 24-28 February 2014, Tokyo, Japan. *J. Cetacean Res. Manage. (Suppl.)* 16:369-409.

Kass, R.E. and Raftery, A.E. 1995. Bayes factors. *Journal of the American Statistical Association* 90: 773-95.

Konishi, K., Hakamada, T., Kiwada, H., Kitakado, T. and Walløe, L. 2014. Decrease in stomach contents in the Antarctic minke whale (*Balaenoptera bonaerensis*) in the Southern Ocean. Paper SC/F14/J14 presented to the JARPA II Special Permit Expert Panel Review Workshop, February 24-28 2014, Tokyo, Japan (unpublished). 37pp. [Paper available from the Office of this Journal].

Konishi, K. and Walløe, L. 2014. Time trends in the storage of energy in the Antarctic minke whales during the JARPA and JARPA II research periods. Paper SC/F14/J13 presented to the JARPA II Special Permit Expert Panel Review Workshop, February 24-28 2014, Tokyo, Japan (unpublished). 21pp. [Paper available from the Office of this Journal]