SC/66a/Rep/10 Rev1

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

International Whaling Commission

INTERNATIONAL
WHALING COMMISSION

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

The Workshop was kindly organised by the National Marine Fisheries Service, with support from the International Whaling Commission, in San Diego 21 May 2015. The list of participants is given as Annex A.

1. INTRODUCTORY ITEMS

1.1 Background

In September 2014, the USA's National Marine Fisheries (NMFS) 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 (California, Oregon and Washington). At the conclusion of that workshop, participants noted that a number of independent species distribution models (SDMs) have been developed using various methods and data sets, particularly for blue whales. They agreed that a collaborative effort to develop formal methods to compare and combine predictions from these models was needed.

Considering that the Scientific Committee of the International Whaling Commission (IWC/SC) would be meeting in San Diego, California, from 22 May to 3 June 2015, participants at the NMFS workshop agreed that it would be worthwhile to hold a one-day workshop as a pre-meeting to the IWC/SC meeting to take advantage of the IWC/SC's broad expertise in modelling issues. Contemporaneously, a specific interest in developing guidelines and recommendations for best practices in SDMs for large whales has recently emerged within the IWC/SC (2014 Report of the Scientific Committee, IWC/65/Rep01, Annexes K1 and T35), an area in which NMFS has considerable expertise. A focused group of experts from NMFS, the IWC/SC and other organizations would be invited to contribute to these discussions, and to help chart a way forward towards generating ensemble SDMs for blue whales.

1.1.1 Objectives

The objective of this Workshop was to convene a group of experts in modelling, statistics, and marine ecology for a one-day event to identify methods to compare and combine model predictions, using existing SDMs for the Eastern North Pacific blue whale as a case study.

Specific objectives included:

a) compiling an initial set of 'established' candidate models for comparison,

b) compiling a list of additional/ancillary data sets (e.g. tracking data, opportunistic observations, acoustic detections, and photo-id and focal-follow surveys) that could be used to develop additional models or for model evaluation, and

c) identifying methodological approaches and requirements for conducting model comparison and ensemble averaging at a future meeting(s) to be determined.

1.1.2 Scientific rationale

For several years, the SDM modelling community has recognized the value of using predictions from a set ('ensemble') of models rather than those from a single model because multimodel weighted averages often yield more robust predictions (Wintle et al. 2003, Johnson and Omland 2004, Araújo and New 2007, Thuiller et al. 2008, Gritti et al. 2013). This concept is now being incorporated into SDMs for several marine top predators (Hobday 2010, Oppel et al. 2012, Renner et al. 2013), including cetaceans (Palacios et al. 2013, Forney et al. 2015). Additionally, model comparison and multimodel inference approaches (Burnham and Anderson 2002, 2004, Claeskens and Hjort 2008) can often result in improved model development by identifying the key assumptions and limitations of single models. Among others, this process is notably used by the Intergovernmental Panel on Climate Change (IPCC) and the climate change modelling community to arrive at 'consensus models' (e.g., Sheffield and Wood 2008, Fordham et al. 2012, Zhang et al. 2015).

For blue whales in the Eastern North Pacific, the recent creation of a number of SDMs makes this a feasible candidate for generating robust predictions using ensemble averaging. It is anticipated that an ensemble-averaged distribution model will have important management applications by showing where blue whales may be more vulnerable to different human activities. This process will lay the groundwork for the future development of models for other large whales off the USA's West Coast – a primary area of interest for NMFS, and more broadly for whale populations worldwide managed by the IWC. Further, the development of a framework for multimodel ensemble averaging that includes analytical techniques such as hierarchical modelling, machine learning, and integrative approaches also opens up the possibility for incorporating ‘less traditional’, but potentially valuable data sources (e.g. acoustic monitoring, satellite tracking, short-term tagging, photo-ID, focal-follow surveys, opportunistic observations, citizen science, etc.) into more robust cetacean SDMs.

1.2 Appointment of Chairs and Rapporteurs
Becker, DeAngelis, Palacios and Redfern were elected co-Chairs. Kelly agreed to act as rapporteur.

1.3 Adoption of Agenda
The adopted agenda is given as Annex B.

2. REVIEW OF SDM MODELLING APPROACHES FOR EASTERN NORTH PACIFIC BLUE WHALES, ASSOCIATED AND AUXILIARY DATA SETS.

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 this section, including a compilation of the characteristics of the models and of the data sets presented in Tables 1 and 2. The geographic area where the ensemble average SDM would be created based on these models is depicted in Figure 1.

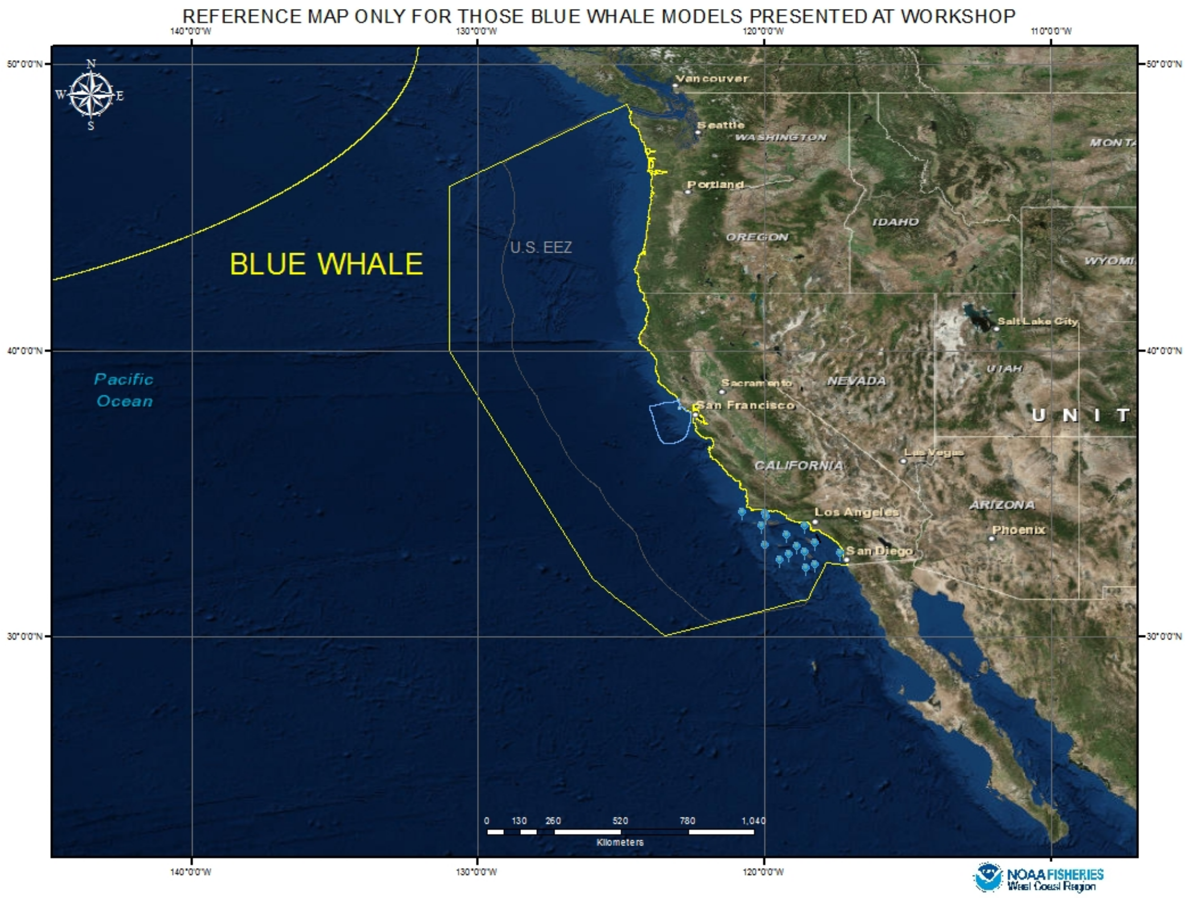

Figure 1. Proposed area off the USA’s West Coast where the ensemble average SDM for the Eastern North Pacific blue whale would be generated (yellow polygon). This area encompasses the domains from the various models that fell within the boundaries for the ensemble. Some of the models encompassed a larger area (e.g. the partial oval at the top left of the map. Some of the model domains were smaller than the proposed area, and they are shown in blue near San Francisco and off Southern California.

2.1 Passive Acoustics
Širović presented work on habitat of calling blue whales in the Southern California Bight (SCB). Eastern North Pacific blue whale "B calls" were detected from passive acoustic data collected over seven years at 16 sites in the SCB. Calling whales were most common in the coastal areas, during the summer and fall months. These data were used to develop habitat models of blue whales in the SCB, using remotely sensed covariates such as sea surface temperature, sea surface height, chlorophyll-*a* and primary productivity. A random forest framework was used for variable selection, and generalized additive models were developed to explain functional relationships and to investigate the predictive abilities of models of calling blue whales. Seasonal component was an important feature of all models, but in areas of high calling abundance there was also a negative relationship with the sea surface temperature. In areas of lower abundance, the chlorophyll-*a* concentration was an important predictor variable. Predictive models generally performed better for general trends than absolute values, but there was a large degree of variation in year-to-year predictability across different sites.

Monnahan described the use of a beta-binomial GAMLSS model to acoustic data from hydrophones across the North Pacific to estimate the broad spatiotemporal distribution of Eastern North Pacific blue whales. The model predicts the probability of observing a song call type from a male for a given location and month. Position and month were the only predictors used, because other predictors, such as oceanographic or biological, were unavailable for catches. The presence/absence acoustic data demonstrated complex patterns of over-dispersion, so an additive form was also used for the over-dispersion term of the beta-binomial model. Uncertainty in model fits was estimated using non-parametric bootstrapping. This model was combined with a similar one for the Western North Pacific acoustic data, and these were used in conjunction to estimate how many historical catches were from the eastern population.

2.2 Line-transect data
Forney presented various habitat-based models of cetacean density that were developed and validated for the Central North Pacific. Spatial predictions of cetacean densities and measures of uncertainty were derived based on data collected during 15 large-scale shipboard cetacean and ecosystem assessment surveys conducted from 1997 to 2012. Generalized additive models of cetacean density for 10 species were developed using static and remotely sensed dynamic habitat variables, including distance to land, sea-surface temperature (SST), standard deviation of SST, surface chlorophyll-*a* concentration (CHL), sea surface height (SSH) and SSH root-mean-square variation. Four basic models were developed: including or excluding latitude and longitude, and including or excluding CHL. Latitude and longitude can help explain spatial patterns but also may prevent models from being sufficiently dynamic to account for temporal habitat variation. CHL has been shown to be an important predictor, but was not available for the 1997 survey, so there was a trade-off between losing the predictor CHL and losing the only survey year covering the northeastern portion of the study area. Predictors for each model type were selected using AIC, and density predictions were averaged across monthly predictor grids within the study period. For all species, the four candidate models exhibited similar performance based on explained deviance, prediction error sum-of-squares and observed:predicted ratios, and there was no single 'best' model. Therefore, a discrete model-averaging approach was implemented with all four models weighted equally to produce final model-averaged density predictions. Uncertainty was estimated using a jackknife procedure in which 10% of the survey days were excluded and the entire model selection and model averaging approach was repeated to estimate model uncertainty as well as temporal variation within the study period. Survey days were chosen as the jackknife sampling unit rather than the smaller individual samples, because survey days provided a greater degree of independence between samples, which is important for characterizing variation adequately. The model-averaging approach allowed identification of areas where uncertainty in the predicted densities was greater (i.e. in unsampled or poorly sampled areas) or lesser (i.e. in areas of greater survey coverage). This study demonstrated that model-averaging, with jackknife uncertainty estimation, is computationally feasible within the GAM framework. Although the four models in this study were weighted equally, alternate weighting schemes could be considered if objective measures can be applied.

Becker described another line-transect survey-based SDM for blue whales in the Eastern North Pacific Ocean. Data collected from Southwest Fisheries Science Center (SWFSC) line-transect surveys conducted in the California Current Ecosystem (CCE) in the summer and fall, 1991-2009, were used to build habitat-based density models for blue whales to support Navy environmental planning activities. Candidate models were built with two sources of dynamic habitat data: (1) remotely sensed and oceanic variables collected *in situ* during the surveys, and (2) modelled oceanic data from the Regional Ocean Modelling System (ROMS). Samples for modelling were created by dividing the continuous survey effort into segments of approximately 5-km length. Generalized additive models (GAMS) were developed in the R package 'mgcv' using restricted maximum likelihood (REML) to optimize the parameter estimates. The number of individuals detected on each segment was modelled as the response variable, using a Tweedie error distribution to account for overdispersion. The natural log of the effective area searched was included as an offset, and incorporated segment-specific estimates of the effective half-strip width (ESW) and the probability of detecting a group of animals on the trackline [*g*(0)]. Density predictions for distinct 8-day composites covering the survey periods were averaged to produce spatial grids of blue whale density at 10-km resolution within the CCE study area, as well as spatially explicit measures of uncertainty. Model performance was assessed using the percentage of explained deviance, root-mean-squared error, observed to predicted density ratios, and visual inspection of predicted and observed distributions. In addition, model-based abundance estimates for the study area were compared to separately derived line-transect estimates. The best model included ocean depth and dynamic ROMS predictors: sea surface temperature, sea surface height, and an interaction term of mixed layer depth and latitude. Predicted distributions showed good concordance with observed sightings and Biologically Important Areas identified for blue whale using an independent dataset. Study area abundance estimates were similar to those derived from standard line-transect analyses of the same data. The SWFSC blue whale habitat model was developed from a set of repeated systematic survey data spanning almost 20 years, and these data provide an important contribution~~brings~~ to the ensemble model ~~representation of blue whale distributions a set of repeated systematic survey data spanning almost 20 years~~. It also ensures that the ensemble model captures broad-scale patterns of blue whale distribution offshore. However, given the sampling resolution of the SWFSC surveys, model predictions may not capture fine-scale distributions patterns, particularly near the coast.

A third study using line-transect data to produce blue whale SDMs was described by Jahncke. Environmental factors determine abundance and distribution of blue whales (*Balaenoptera musculus*) and their prey (*Euphausia pacifica and Thysanoessa spinifera*) within Gulf of the Farallones and Cordell Bank National Marine Sanctuaries. The purpose of this study was to identify potential areas of high probability of blue whale and krill co-occurrence, and discuss the management implications of existing shipping regulations in relation to the locations and temporal patterns of these predictable foraging ‘hotspots’. Blue whale data were collected along ACCESS cruises from the vessel’s flying bridge using standardized methods while ‘on effort’ at 10 kt. The total number of blue whales counted along each transect line was summed for each 3-km bin and assigned to each bin’s midpoint (*n*=4068; 3986 zero-count bins, 82 non-zero-count bins). Separate predictive models were developed of whale abundance (using negative binomial regression on count data) and krill abundance (using a two-part model combining logistic and negative binomial regressions on acoustic biomass) over a 10-yr period (2004–2013). Variables were selected assuming bottom-up regulation and that blue whales (and krill) aggregate at predictable locations where the interaction between oceanography and bathymetry enhances foraging opportunities. Variables included were in-situ surface and mid-water oceanographic measures (temperature, salinity, and fluorescence), regional- and basin-scale climate indices (UI, SOI, PDO, NPGO), and bathymetric- and distance-related data. Univariate testing was used to select the most significant of the proper transformations of variables to be included in first full model. The negative binomial model with manual backwards stepwise removal was used until all variables in the model were significant ($p < 0.05$). P-values were determined using the Likelihood Ratio Test (LRT). A subset of environmental variables was considered with respect to interaction with year. Interactions (up to two) were retained in a model if they were significant for the LRT. Finally, models were evaluated for multi-collinearity using the Variance Inflation Factor, and model validation was conducted using k-fold cross validation (k = 10; 20 simulations). Predictions were applied on a 1-km^2 cell prediction matrix spanning the Sanctuaries for May, June, July, and September as well as for each of the 10 yrs of surveys to identify persistent distribution patterns. Highest-use areas were explored. Additionally, the footprint of the San Francisco Bay shipping lanes was overlaid to assess where there is high human use (shipping) and important foraging areas (overlap of blue whales and krill predator). Both whales and krill were found to consistently use the northeast region of Cordell Bank, the Farallon Escarpment and the shelf-break waters. All identified blue whale hotspots were also krill hotspots, indicating the value of identification of important areas of prey concentration to assist management of blue whales. Areas north of Cordell Bank are of particular concern as they overlap with the end of the San Francisco Bay northern shipping lane. The fine scale and long-term coverage of the ACCESS data provides strength to the analysis; however, it was limited by few blue whale sightings in a relatively small geographic location. These findings support requests for additional management measures to decrease the threat to whales, particularly in important foraging areas, by implementing vessel speed restrictions in management areas on a seasonal basis.

Pardo described a final blue whale SDM based on line-transect data using a hierarchical Bayesian modelling approach. These hierarchical models were intended for predicting the spatial distribution of blue whales (*Balaenoptera musculus*) from the water-column structure of the Eastern North Pacific, as indexed by the absolute dynamic topography of the ocean’s surface (ADT). This would help in the identification of priority dynamic areas for the species. Predictions are provided as long-term averages of daily-inferred population densities during summer-autumn, spanning 1993-2009. Interannual variations were evaluated for the spatial distribution for years with some sample effort within the study area. Only absolute dynamic topography of the ocean’s surface was used because it is very sensitive to processes of pycnocline shoaling and outcropping that give shape to the most productive habitats of the Eastern North Pacific where blue whales feed actively on low-trophic-level prey. The whale data come from line transect surveys conducted by NMFS-NOAA, which follow distance sampling techniques. The ADT was collected by satellite altimeters and obtained from AVISO. Both databases had a spatial resolution of 0.25-degree cells and were paired on a daily basis from 1993 to 2009. The detection probability sub-model for estimating the population densities was fitted from all sighting distances collected by the NMFS-NOAA across the Eastern North Pacific Ocean from 1986, whereas ADT data were available for fitting the ecological sub-model only from 1993. Two polynomial functions and a mixture of the two were tested for the species-habitat process and the best was chosen based on the lowest Deviance Information Criterion. The algebraic formulation of the model was written explicitly using the JAGS language within R statistical analysis package. Both the observational and the ecological processes were included as sub-models, connected by common parameters in a hierarchical structure. The uncertainty was propagated by the posterior distributions for all parameters (i.e. Bayesian analysis of the model), which were approximated by a Markov Chain Monte Carlo procedure using the Gibbs Sampler algorithm. The prior distributions were assumed to be non-informative, except for the probability of detection on the transect line (i.e. *g(0)*).

2.3 Satellite tag data

Management of highly migratory species requires spatially explicit information on their distribution and abundance, and how these vary over time. Satellite telemetry provides time-series information on individual movements, but these ‘presence-only’ data have generally been ranked lower on management’s data hierarchy than survey-based density estimates. Hazen presented a method for using satellite-tag data to develop real-time spatial management approaches (weeks), specifically estimating density of endangered blue whales (*Balaenoptera musculus*) for use in establishing dynamic management areas. A state-space model was applied to 104 blue whale satellite tracks from 1994 to 2008 to account for errors in the locations. Daily positions were output, and these integrated with remotely sensed environmental variables as a ‘proxy’ (i.e. variables that might not directly, or even indirectly, influence a species distribution, but that are correlated with some factor that does) for habitat preference and processes driving prey availability. ‘Absence’ points were created using a correlated random walk to sample the environment where whales could have gone at the same temporal and spatial scales as the tag data. Generalized additive mixed and boosted regression tree modelling frameworks were implemented to determine the probability of occurrence, and then scaled by the population abundance to estimate absolute densities which were comparable to survey-based estimated densities. Ultimately, combining survey and tag based model predictions for marine mammals could improve accuracy in data-poor seasons or areas.

Friedlaender explained that since 2010, multi-sensor archival acoustic recording tags have been deployed on blue whales (n=77) in Southern California as part of the SoCal Behavioural Response Study. The primary goals of this project are to test the behavioural responses of marine mammals to sound. However, coincident to that, the collection of an appreciable amount of information on the foraging behaviour and ecology of blue whales has greatly increased knowledge of the species. Thus, the project has recently begun to include quantitative measures of prey in the area around tagged and feeding blue whales (n=34) to determine the ecological relationships between whales and prey and how this drives their observed behaviour. Because of the fine spatial and temporal resolution of the tag data, individual feeding events can be quantified, and these linked to metrics of prey to better understand the scale at which these whales feed, foraging rates and the duration of foraging bouts. Likewise, focal follow information collected throughout tag deployments can be used to measure area-restricted search at the scale of a feeding dive or a foraging bout. Incorporating this information to validate data collected over broad spatial and temporal scales can augment efforts to accurately model the behaviour and ecology of blue whales.

2.4 Associated research examples

To provide a picture of further examples of large cetacean SDMs, Murase presented results of application of different types of SDMs to sei whales (*Balaenoptera borealis*) in the Western North Pacific. A total of eight models [two regression models (GLM, and GAM), three profile models (BIOCLIM, DOMAIN and ENFA) and three machine learning models (BRT, RF and MaxEnt)] were used in the analysis. GLM, GAM, BRT and RF models used presence/absence data while BIOCLIM, DOMAIN, ENFA and MaxEnt used presence only data. The modelling results were evaluated based on the AUC (area under the curve). It was difficult to determine the best estimated spatial distribution maps based entirely on such an index because the resultant maps were highly variable, even if difference of the AUC was subtlely small. Contributions of explanatory variables to the models varied from model to model. The results of this analysis supported previous knowledge that different models would create different prediction results. Ensemble modelling can be one of the solutions to overcome the limitations of a single model if the main aim of a study is prediction of spatial distribution of target species. However, it might be difficult to achieve ecological inference (e.g, niche and habitat) from the results of ensemble modelling. Because contributions of explanatory variables were not consistent, ecological inference from SDMs is still open to question.

2.5 General discussion about the presentations

Whilst prey-field data are considered to be a likely influential covariate for describing the distribution of baleen whales on their feeding grounds, it has remained difficult to measure, or even infer, these in a synoptic sense, particularly across ocean-basin scales. Aggregations of species such as krill can be ephemeral, even at local distances and daily time-scales. Chlorophyll-*a* and primary productivity measures may sometimes scale well with krill distribution, but other times not, so these are accessible but imperfect approximations for the prey field.

In terms of additional prey-related datasets, there are various mid- to low-trophic level surveys occurring in the East North Pacific, including the CalCOFI ship surveys and the aerial surveys for sardines and also the NMFS rockfish surveys in May-June, which co-incidentally measure mesozooplankton (i.e. krill) with active acoustics (i.e., as described in Santora et al. 2011). However, given the timing and location of these surveys, they are unlikely to overlap with blue whale presence. Other surveys may exist that could provide useful prey-field data, and the Workshop **agreed** that attempts to link to these should be made.

There was some discussion about potentially useful 'proxy' variables that probably do not directly, or even indirectly, influence whale distribution, but that may correlate with processes that do. The obvious example in the context of Eastern North Pacific blue whales is latitude, which correlates well with processes like day length, but that may also be useful in describing north-south environmental gradients or ecosystem differences. Inclusion of latitude as a predictor variable is likely to be helpful, but, obviously care should be taken if extrapolating for this variable beyond the survey sample space.

3. REVIEW OF METHODOLOGICAL APPROACHES AND REQUIREMENTS FOR MODEL COMPARISON AND ENSEMBLE AVERAGING

3.1 Purpose

What is the purpose of an SDM? ('SDM' is used here as a catch all.) For the purposes of this workshop, the objective of the outputs of SDMs is to inform managers when making decisions that aim to minimize human-induced impacts on species/populations/areas (e.g. with respect to ship strikes). The first step is to begin to derive and define explicit management needs. But the question of how to derive sampling protocols and subsequent modelling is not necessarily straightforward for a number of different reasons. For example, it is often difficult for either managers to specify in a quantitative way or for scientists to determine what is actually required or wanted because of the case-by-case nature of a task and the sometimes multiple threats being addressed. Often managers are not aware of what they might actually ask, beyond desiring the very best scientific foundation to their decisions. There is thus an important need for an iterative process involving both managers and scientists as to what is needed, and what is possible from a science perspective. It is essential that scientists advise on the limitations of models and their outputs; this includes clear statements about assumptions, and estimates of comprehensive errors for models.

Another associated question is whether a given SDM is predictive or explanatory? This dichotomy reflects whether models were produced with the aim of providing precise and accurate predictions strictly within the sampling space the data were collected, or whether there is a desire to describe the ecological mechanisms driving the distribution and abundance of a species of interest. A model formulated with the predictive aim can only predict accurately within the sampling space on which it is based, and will eventually fail if/when underlying processes or relationships change. An explanatory (i.e., mechanism-driven) model is more likely to be robust to predictions outside of that sampling space, but such models often need more parameters, which can lead to less reliable extrapolations. That is, just because a model has a mechanism does not make it better at prediction or extrapolation. Within these bounds, there was consensus that ideally all models would help determine ecological mechanisms, and that SDMs based on explanatory or mechanism-driven formulations are more likely to allow incremental model improvements, as knowledge is tested and accepted.

3.2 Scale

There are multiple interacting scales inherent to SDMs, ranging from ecological scales (at what time and spatial scales are blue whales interacting with their environment), fitting scales (which relate to errors for sampling and scales of oceanographic variables used in the model), and prediction scales (what time and spatial scales are used for prediction). Questions about which spatial and temporal scales are used are fundamental in production of SDMs. Which spatial and temporal scales are used in models is often reliant on individual data streams, but should also be aligned with management requirements; this can often be a point of tension in developing SDMs.

On one hand, it may seem appropriate to consider that the finest scale relevant is the temporal and geographical scale that is relevant to the animal; however, this is rarely known. Cetaceans are highly mobile, and accurately reflecting their movement in space and time in a modelling framework will be extremely complex, as will obtaining data at such a scale for explanatory variables. From a model perspective, there will be a wide range of scales for processes that will be modelled, and there is a wide range of scales from which data will become available – and these are not necessarily congruent. Naturally, it is not trivial to reconcile these potential discrepancies. However, the scales of data upon which models are supported need not restrict the scales at which predictions are made. The important question remains as to whether 'fine scale' (or, finest scale in the context of the model in question) is the best 'default' option for fitting SDMs?

Although it is unlikely that a one-size-fits-all solution can be found for dealing with the question of scale in cetacean SDMs, the following suggestions were offered in discussion.

- Given the different scales of extent and resolution within the CCE, an ensemble approach should consider nesting within spatial and temporal scales, so that processes at different scales can be integrated.
- Sensitivity analyses should be undertaken to determine any effect of predicting with models at different spatial and temporal scales. Furthermore, the effect of changes in scale can be evaluated against implications for management advice, and even be part of model evaluation process.
- Data and/or management questions do not necessarily have to be on the same scales, but there needs to be care on how differences are dealt with, for example, the influence of mismatch needs to be evaluated. In addition, trying to inform management decisions may prove challenging with coarse-resolution models.
- There is potential in the seabird literature (i.e. same ecosystem, but different taxa) to help inform analysis decisions about spatial and temporal scales for cetacean analyses. In particular, spatial and temporal scales make an important difference to seabirds, and, as an extension, will probably make a difference to cetacean modelling.
- Avoid trying to produce 'all embracing' models, to answer ALL questions, at all scales, at once.
- Modellers need to evaluate the limitations of using proxy variables in describing relationships with whale presence or densities, and how these limitations might cascade through varying spatial and temporal scales. This is particularly the case for blue whales or any species that migrates seasonally into an ecosystem such as the CCE, but whose range may include the entire North Pacific.

3.3 Criteria for evaluating specific models

A key motivation for the Workshop was to consider and evaluate methods to produce ensemble model(s), for example as described in Araujo and New (2007), for blue whales in the Eastern North Pacific. A range of models had been presented at the Workshop to describe the presence, distribution and abundance of blue whales, including GAMs, boosted regression trees, MaxEnt, etc., as generated by a range of different data types, such as line-transect, tagging and passive acoustics (Table 1). This range of modelling approaches and data, coupled with potential differences in associated spatial and temporal scales (see discussion above) and how the issues of uncertainty are dealt with, makes the task of objective comparison and/or ranking individual SDMs complex.

In the context of the present Workshop, where prediction to provide management advice is a key component, the challenge is to determine the pros and cons of each model and to use that information to estimate 'weights' for each model under an ensemble model framework. Ideally, any weighting approach will be quantitative and objective, and reflect the predictive ability of each model in the light of management objective(s). An important consideration may also be to evaluate seasonal or spatial weighting schemes to allow a model that, say, performs particularly well for inshore habitat in the winter, to have a larger influence within the ensemble for that period. It is also important to consider whether models have been developed to account for, and to aggregate across, major sources of error during their development?

There was also discussion about exploring the performance of models, with particular reference to robustness across different models (e.g. when/where they fail and when/where they appear relatively robust). In terms of comparison amongst models factors to consider include under what circumstances are the predictions similar or different, how comparable are the uncertainty estimates, etc. Here, one would be seeking support for results by exploring agreements in predictions from models being compared, and a lack of support in the presence of discrepancies. It is important to understand how particular models fail under particular circumstances.

The Workshop recognised that standard model-selection statistics (for example, but not limited to, AIC), may not be the appropriate approach. In the application of those statistics, the same data must be used to develop the models that are to be tested. There was also discussion around the use of validation data in assessing model performance, and whether ancillary data, or an independent/new dataset (e.g. where 'truth' is known) should be used for this purpose.

3.3.1 Recommendations and conclusions

The Workshop **agreed** that there is probably a reasonably mature body of literature on model averaging (e.g. AIC 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 appropriate approach for use within the present blue whale situation.

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. For instance, suppose there are two models (e.g. GAM, MaxENT), with 1 and 20 variants respectively. If these 21 models are used with equal weights in the ensemble, the second model will clearly dominate the result. Thus, 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 (as in Tables 1 and 2 of this report) for each candidate model for an ensemble is necessary. The metadata would contain information on key management questions; spatial and temporal scales; how error is estimated and propagated, and whether correlation structure of errors has 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 for 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.

3.2 Specific techniques for developing an ensemble model

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; see Araujo and New, 2007). The Bayesian framework, using the existing predictive output, was also considered, but further discussion on that and other methods is required. However, regardless of the method ultimately used, combinations and comparisons of models must be made on identical units (i.e. whether relative or absolute densities, probabilities of occurrence, etc.) and/or management implications.

4. WORK PLAN AND BUDGET CONSIDERATIONS

The Steering Group for this workshop (Becker, DeAngelis, Palacios and Redfern) was appointed to continue to advance the objectives set out for this workshop. Kitakado was also appointed to the Steering Group. To facilitate progress, the Workshop identified four sequential phases, as follows:

Phase 1

Form three correspondence groups (CG) with the membership to be confirmed in the weeks following conclusion of this Workshop to conduct the following tasks:

CG 1: Review statistical literature and report on techniques that are available and appropriate for building ensemble models.

Membership still needs to be determined, but the following skills are needed: experience in building ensemble models (this expertise should come from outside the usual Scientific Committee membership e.g. a climate modeller), numerical modelling (e.g. a nested ROMS modeller), habitat modelling, and general statistical knowledge.

CG 2: Conduct comparable sensitivity analyses to understand the effect of, inter alia, prediction scale for models to be considered for the ensemble.

Membership should include at least the invited researchers who presented models relevant to ensemble averaging at this Workshop. This group should consider a time frame and budgetary needs to conduct this work.

CG 3: Identify potential management objectives/advice required (including whether there is a need for real-time predictions vs. long-term averages).

Membership should include individuals from NOAA Fisheries' National Marine Fisheries Service and National Marine Sanctuaries, U.S. Navy, U.S. Bureau of Ocean and Energy Management, and the IWC SC.

Phase 2

CG members must complete their projects by February 2016. This may require consultations via teleconference as well as email. In February, the products from the CGs will be presented to and reviewed by individuals from all CGs and other invited participants via a conference call. During this call, plans will be developed for finalizing the products by incorporating reviewer comments. Individuals will also be identified to create a draft plan (i.e., a 'straw man') for developing the ensemble model.

Phase 3

CG members must have final products and a draft plan for developing an ensemble model complete by May 2016. In May, a workshop will be convened to share the products and plan to members of the joint NMFS-IWC May 2015 workshop and other invited participants. This workshop will be either a pre-IWC meeting or a separate workshop, depending on budgetary considerations. At this meeting, plans for developing the ensemble model will be finalized and budget needs for conducting the ensemble modelling work defined. Results of this workshop will be presented at the 2016 IWC SC meeting.

Phase 4

The ensemble model will be developed. This model will be presented at the 2017 IWC meeting.

Budget considerations: Funding at a similar level as required to conduct this workshop will be required to conduct the 2016 intersessional workshop identified in Phase 3. 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 straw man plan developed during Phase 2.

5. ADOPTION OF THE REPORT

The report was adopted by correspondence on 23 May 2015. The co-Chairs thanked all of the participants for their co-operative approach to the Workshop and for the careful review of new and existing information. The Workshop thanked the NMFS and IWC for the financial and logistical support. The Workshop also thanked the co-Chairs for their efficient handling of the discussions and Kelly for her invaluable rapporteuring assistance.

REFERENCES

Araújo MB, New M (2007) Ensemble forecasting of species distributions. Trends in Ecology and Evolution 22:42–47. http://dx.doi.org/10.1016/j.tree.2006.09.010

Becker, E.A., Foley, D.G., Forney, K.A., Barlow, J., Redfern, J.V., Gentemann, C.L., (2012) Forecasting cetacean abundance patterns to enhance management decisions. Endangered Species Research 16, 97–112. http://dx.doi.org/10.3354/esr00390

Burnham KP, Anderson DR (1998) Model selection and inference. A practical information theoretic approach. Springer-Verlag, New York, NY

Burnham KP, Anderson DR (2002) Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. 2nd edition Springer-Verlag, New York.

Burnham KP, Anderson DR (2004) Multimodel inference understanding AIC and BIC in model selection. Sociological Methods & Research 33(2): 261-304, http://dx.doi.org/10.1177/0049124104268644

Calambokidis, J., Steiger, G.H., Curtice, C., Harrison, J., Ferguson, M.C., Becker, E., DeAngelis, M., Van Parijs, S.M. (2015) 4. Biologically Important Areas for Selected Cetaceans Within U.S. Waters – West Coast Region. Aquatic Mammals 41, 39–53. http://dx.doi.org/10.1578/AM.41.1.2015.39

Claeskens G, Hjort NL (2008) Model selection and model averaging. Cambridge University Press, New York, NY

Fordham DA, Wigley TML, Watts MJ, Brook BW (2012) Strengthening forecasts of climate change impacts with multi-model ensemble averaged projections using MAGICC/SCENGEN 5.3. Ecography 35: 4–8. http://dx.doi.org/10.1111/j.1600-0587.2011.07398.x

Forney KA, Becker EA, Foley DG, Barlow J, Oleson EM (2015). Habitat-based models of cetacean density and distribution in the central North Pacific. Endangered Species Research 27: 1–20. http://dx.doi.org/10.3354/esr00632

Gritti ES, Gaucherel C, Crespo-Perez M-V, Chuine I. (2013) How can model comparison help improving species distribution models? PLoS One, 8, e68823. http://dx.doi.org/10.1371/journal.pone.0068823

Hobday, A.J., (2010) Ensemble analysis of the future distribution of large pelagic fishes off Australia. Progress in Oceanography 86, 291–301. http://dx.doi.org/10.1016/j.pocean.2010.04.023

Johnson JB, Omland KS (2004) Model selection in ecology and evolution. Trends in Ecology and Evolution 19:101–108. http://dx.doi.org/10.1016/j.tree.2003.10.013

Monnahan, C.C., Branch, T.A., Stafford, K.M., Ivashchenko, Y.V., Oleson, E.M. (2014) Estimating Historical Eastern North Pacific Blue Whale Catches Using Spatial Calling Patterns. PLoS ONE 9, e98974. http://dx.doi.org/10.1371/journal.pone.0098974.t006

Murase, H., Friedlaender, A., Kelly, N., Palacios, D.M., Palka, D. (2015) A preliminaly review of species distribution models (SDMs) applied to baleen whales. Progress report of the intersessional corresponding group 'Applications of species distribution models (SDMs)'. SC/66a/EM/3. Paper submitted to the 2015 IWC Scientific Committee Meeting, San Diego, California, USA, 22 May – 3 June 2015. 17 pp.

Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O'Connell, A.F., Miller, P.I., Louzao, M., (2012) Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. Biological Conservation 156, 94–104. http://dx.doi.org/10.1016/j.biocon.2011.11.013

Palacios DM, Baumgartner MF, Laidre KL, Gregr EJ (2013) Beyond correlation: integrating environmentally and behaviourally mediated processes in models of marine mammal distributions. Endangered Species Research 22:191-203, http://dx.doi.org/10.3354/esr00558

Pardo, M.A., Gerrodette, T., Beier, E., Gendron, D., Forney, K.A., Chivers, S.J., Barlow, J., Palacios, D.M. (2015) Inferring Cetacean Population Densities from the Absolute Dynamic Topography of the Ocean in a Hierarchical Bayesian Framework. PLoS ONE 10, e0120727. http://dx.doi.org/10.1371/journal.pone.0120727.s002

Renner, M., Parrish, J.K., Piatt, J.F., Kuletz, K.J., Edwards, A.E., Hunt, G.L., Jr, (2013) Modeled distribution and abundance of a pelagic seabird reveal trends in relation to fisheries. Marine Ecology Progress Series 484, 259–277. http://dx.doi.org/10.3354/meps10347

Santora, J.A. et al. (2011). Mesoscale structure and oceanographic determinants of krill hotspots in the California Current: Implications for tropic transfer and conservation. Progress in Oceanography, 91: 397-409.

Sheffield J, Wood EF (2008) Projected changes in drought occurrence under future global warming from multi-model, multi-scenario, IPCC AR4 simulations. Climate Dynamics 31(1): 79-105, http://dx.doi.org/10.1007/s00382-007-0340-z

Thuiller W, Lafourcade B, Engler R, Araújo MB (2009) BIOMOD – a platform for ensemble forecasting of species distributions. Ecography 32: 369-373, http://dx.doi.org/10.1111/j.1600-0587.2008.05742.x

Wintle BA, McCarthy MA, Volinsky CT, Kavanagh RP (2003) The use of Bayesian model averaging to better represent uncertainty in ecological models. Conservation Biology 17:1579−1590. http://dx.doi.org/10.1111/j.1523-1739.2003.00614.x

Zhang L, Liu S, Sun P, Wang T, Wang G, et al. (2015) Consensus Forecasting of Species Distributions: The Effects of Niche Model Performance and Niche Properties. PLoS ONE 10(3): e0120056. http://dx.doi.org/10.1371/journal.pone.0120056

Table 1.

Compilation of the characteristics of the models presented at the Workshop.

Name	Data Source	Purpose	Response Variable	Prediction Scale	Habitat Variable	Region	Method	Time Period	Candidate for ensemble?
Širović	Acoustic	Navy impact assessment	Call density	50km/recorder	SST, SSH, Chlorophyll-*a*, Primary productivity	Southern California only	Random Forests and GAM	Year-round (2006-2012)	Validation
Becker	SWFSC line transect [CCE]	Abundance and Distribution	Density	10-km grid	SST, SSH, MLD, depth	West Coast	GAM	Summer and fall (1991-2009)	Yes
Pardo	SWFSC line transect [CCE]	Exploring mechanism-driven hypothesis	Density	0.25-deg grid	ADT	West Coast	Hierarchical Bayesian	Summer and fall (1993-2009)	Yes
Jahncke	ACCESSS line transect	Probability of encounter	Blue whale counts	3-km bins	PDO-1, NPGO-3, contour index, distance to mainland, shelf break	Fine-scale limited to San Francisco area	Negative binomial regression	2004-2013	Yes – for a small part of the CCE
Hazen	Tagging	Density	Real whale location	25-km grid	SST, CHL, SSH, meridional winds, depth, slope,	West Coast	BRT and GAMM	1994-2008	Yes

					aspect, distance to shelf				
Monnahan	Acoustic	Population identity	Prob. of observing >1 one call/hr	Continuous coordinate surface	Latitude, longitude, month	Whole north Pacific	GAMLSS	1992 - 2004	No
Friedlaender	Tagging	Measure behavioural change in cetaceans from sound stimulus				Southern California Bight		2010 - 2014	Validation

Table 2

Compilation of other sources of data identified at the Workshop (incomplete list of all sources available)

Name	Data Source	Purpose	Response Variable	Prediction Scale	Habitat Variable	Region	Method	Time Period	Candidate for ensemble?
Irvine	Tagging	Number of overlapping home ranges				West Coast			Yes, but requires special considerations and further work
Cascadia Research Collective, Inc.	Photo id and tagging	Areas and timing of large aggregations and behaviour				West Coast			No – validation

Annex A
List of Participants

Name	Affiliation
Elizabeth Becker (co-convenor)	National Marine Fisheries Service (USA)
Monica DeAngelis (co-convenor)	National Marine Fisheries Service (USA)
Daniel Palacios (co-convenor)	Oregon State University (USA)
Jessica Redfern (co-convenor)	National Marine Fisheries Service (USA)
Natalie Kelly (rapporteur)	CSIRO (Australia)
Debra Palka	National Marine Fisheries Service (USA)
Charlotte Boyd	National Marine Fisheries Service (USA)
Doug Butterworth	University of Cape Town (South Africa)
William de la Mare	Australian Antarctic Division (Australia)
Greg Donovan	International Whaling Commission
Samba Diallo	COMHAFAT (Guinea)
Mike Double	Australian Antarctic Division (Australia)
Paul Fiedler	National Marine Fisheries Service (USA)
Karin Forney	National Marine Fisheries Service (USA)
Ari Friedlaender	Oregon State University (USA)
Caterina Fortuna	ISPRA (Italy) & IWC
Tim Gerrodette	National Marine Fisheries Service (USA)
Anita Gilles	National Marine Fisheries Service (Alexander von Humboldt Foundation Fellow) (USA)
Elliott Hazen	National Marine Fisheries Service (USA)
Jaime Jahncke	Point Blue Conservation Science (USA)
Suzanne Manugian	Point Blue Conservation Science (USA)
Cole Monnahan	University of Washington (USA)
Hiroto Murase	National Research Institute of Far Seas Fisheries (Japan)
Konan Nda	COMHAFAT (Cote d'Ivoire)
Mario Pardo	CICESE (Mexico)
Doug Sigourney	National Marine Fisheries Service (USA)
Ana Širović	Scripps Institution of Oceanography (USA)

Annex B

Agenda

1. Introductory items

    1.1 Convenors' 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 of SDM modelling approaches

3. Review of data sets available for SDMs

4. Review of methodological approaches and requirements for model comparison and ensemble averaging

5. Work plan and budget considerations

6. Adoption of report