SC/67A/EM/15

An update of review of species distribution models (SDMs) applied to baleen whales and a guideline on the techniques and underlying assumptions of the models

Hiroto Murase, Ari Friedlaender, Natalie Kelly, Toshihide Kitakado, John Mckinlay, Daniel M. Palacios and Debra Palka

INTERNATIONAL
WHALING COMMISSION

An update of review of species distribution models (SDMs) applied to baleen whales and a guideline on the techniques and underlying assumptions of the models

(Progress report of the intersessional corresponding group "Applications of species distribution models (SDMs)" in the period between IWC/SC 66b and 67a)

HIROTO MURASE[1], ARI FRIEDLAENDER[2], NATALIE KELLY[3], TOSHIHIDE KITAKADO[4], JOHN MCKINLAY[3], DANIEL M. PALACIOS[2] AND DEBRA PALKA[5]

[1] *National Research Institute of Far Seas Fisheries, Japan Fisheries Research and Education Agency, 2-12-4 Fukuura, Kanazawa, Yokohama, Kanagawa 236-8648, Japan*
[2] *Marine Mammal Institute, Department of Fisheries and Wildlife, Hatfield Marine Science Center, Oregon State University, 3020 Marine Science Drive, Newport, OR 97365 USA.*
[3]*Australian Antarctic Division, 203 Channel Highway, Kingston, Tasmania 7050, Australia*
[4]*Tokyo University of Marine Science and Technology, 4-5-7 Konan, Minato-ku, Tokyo, 108-8477, Japan*
[5]*NOAA Fisheries, Northeast Fisheries Science Center, 166 Water Street, Woods Hole, Massachusetts, 02543, USA*

ABSTRACT

An intersessional correspondence group has been established since IWC/SC65b to contribute development of a guideline on the techniques and underlying assumptions of SDMs (species distribution models) based on up-to-date and comprehensive knowledge (IWC, 2015). The group conducted a preliminary review of SDMs applied to baleen whales (Murase et al., 2015) and preliminary reviews of machine learning methods which are commonly used as SDMs (Murase et al., 2016). During the intersessional period from IWC/SC 66b to 67a, the group updates Murase et al. (2015) adding information from Murase et al. (2016) as well as 12 peer-reviewed papers published after March 2015. A total of 48 papers are now considered in the updated review. The group intends to finish the work during the intersessional period from IWC/SC 67a to 67b.

INTRODUCTION

The intersessional correspondence group "Applications of species distribution models (SDMs)" has been established by the EM working group since IWC/SC 65b (IWC, 2015). Terms of reference of the intersessional corresponding group ("SDM group" hereafter) are to "develop guidelines and recommendations for best practice in species distribution modelling". In the first year, the SDM group conducted a preliminary review of SDMs applied to baleen whales and submitted the results to IWC/SC 66a (Murase et al., 2015). In the second year, the SDM group conducted preliminary reviews of machine learning methods applied as SDM and submitted the results to IWCSC/66b (Mursae et al., 2016). During the intersessional period from IWC/SC 66b to 67a, the SDM group updated Murase et al. (2015), adding information from Murase et al. (2016) as well as 12 peer-reviewed papers published after March 2015. This paper presents the results of the update.

Within the Scientific Committee of the International Whaling Commission (IWC/SC), a generalized additive model (GAM) based SDM was developed in the late 1990s for generating an abundance estimate for Antarctic minke whales (*Balaenoptera bonaerensis*) (Hedley *et al.*, 1999). Traditionally, abundance of baleen whales for the purpose of management under the Revised Management Procedure (RMP) has been estimated using statistical design-based methods, such as distance sampling (Thomas *et al.*, 2010), according to a guideline from the IWC/SC (IWC, 2012). The Sub-Committee on the RMP of the IWC/SC is currently trying to develop a guideline for model-based abundance estimation methods, mainly focusing on GAMs and a pre-meeting is to be held prior to IWC/SC 67a to test the proposed new guideline against test cases.

Together with the review conducted by the SDM group and the outcome from the pre-meeting on the proposed new guideline for model-based abundance estimation methods, the SDM group plans to complete the task assigned to the group during the intersessional period from IWC/SC 67a to 67b.

GENERAL REVIEW OF SDMS

Background

Spatial distribution of biological organisms is one of the fundamental information sources used for management and conservation of the target species. Classically, geographic locations of target species' occurrence are plotted on maps to understand their spatial distributions. Application of species distribution models (SDMs) has proliferated since the 1990s in parallel with the advancement of computing power, software such as geographic information systems (GIS) and statistical techniques (Palacios *et al.*, 2013). Application of SDMs to baleen whales started in the late 1990s and the number of studies has been increasing in recent years. In this manuscript, a statistical model relating occurrence of a species to its environment within a certain time period is termed an SDM. Such a model can be used to predict spatial distributions of the target species once the model is constructed. An SDM is not a mechanistic model that can deal with dynamic processes of the spatial distributions but an empirical model that can incorporate observed relationships between occurrence of species and their environment at a certain time period (Palacios *et al.*, 2013). A SDM can also be called as a snapshot model, being static rather than dynamic.

Other terms, such as habitat distribution model (Guisan and Zimmermann, 2000), ecological niche model (Peterson *et al.*, 2011) and resource selection function (Manly *et al.*, 2002), are sometimes used interchangeably with SDM. Habitat, niche and resource selection modelling are important and interesting topics in the field of ecology, but the definition of these terms can be bewildering and the interpretation of output from models in these contexts is sometimes difficult, especially for baleen whales as they can use a large geographic area and display complex behaviour such as migration between breeding and feeding grounds. A GAM-based spatial model estimating abundance which has been corrected for uncertain detection via the distance sampling method is specifically called a density surface model (DSM) (Miller et al., 2013). The term "SDM" is used throughout this manuscript because the main focus is prediction of spatial distribution of target species. However, it should be noted that within the SDM framework, there is also scope for addressing ecological questions and hypotheses, but one does have to be careful about the selection and interpretation of explanatory and/predictor variables. Detailed textbooks dealing with these kinds of models are available, such as Franklin (2009) and Peterson *et al.* (2011), and these are referred throughout this manuscript.

A review paper of habitat modelling techniques for cetaceans was published in *Marine Ecology Progress Series* in 2006 (Redfern et al., 2006), and a later review of the application of SDMs to marine species was also published by Robinson *et al.* (2011). A series of workshops on habitat modelling for marine mammals have been held at the Biennial Conference of the Society for Marine Mammalogy since 2001, and these activities led to the theme section in the journal *Endangered Species Research* titled "Beyond Marine Mammal Habitat Modeling" (Gregr et al., 2013). Several papers in that theme section are reviewed in this manuscript.

The aim of this review is threefold. First, general aspects of SDMs are briefly reviewed. Second, applications of SDMs on baleen whales are reviewed. To accomplish the second aim, a total of 48 published papers are reviewed. Finally, preliminary recommendations are provided to develop practical guidelines. Some habitat modelling studies did not predict spatial distributions, even if similar statistical models to SDMs were used (e.g., Friedlaender et al., 2006). Habitat modeling studies without prediction of spatial distribution are not considered fully in this manuscript. Similarly, geostatistical methods such as inverse distance weighting and kriging have also been used to predict spatial distributions of baleen whales. However, geostatistical methods are not reviewed fully in this manuscript because, with this method, spatial distribution of species are predicted based on geographic coordinates without consideration of environmental factors/drivers.

Data for SDMs

Data on occurrence/abundance of species (response variable)

To construct SDMs, data on geographic locations of observed species (i.e., response variable) and environmental data at the time of observations (i.e., explanatory or predictor variables) is required. Geographic locations and timestamps of the observations are fundamental data for SDMs. There are two basic types of response variables: (i) occurrence (when and where individuals and groups were detected during a survey) and (ii) abundance (i.e., number of individuals detected). Abundance (or, equally, density) data can be considered an extension of occurrence data. Probability of occurrence can be estimated if occurrence data are used, while abundance can be estimated if abundance data are used.

If occurrence or abundance data are obtained from a target area based on survey activity recording when and where search effort was undertaken (regardless of whether such a survey is based on a predetermined statistical survey design, or an opportunistic, non-random design), both occurrence (i.e. presence) and absence data can be obtained. In contrast, data such as telemetry and catch-without-effort data can only provide occurrence data. SDMs can be categorized by four approaches based on available response variables: (i) presence/absence, (ii) presence only, (iii) presence/background, and (iv) presence/pseudo-absence. Background data implies the existence of co-collected environmental data in a target area. Pseudo-absence implies existence of environmental data in a target area, but environmental data at occurrence positions are excluded. Presence/background and presence/pseudo-absence approaches were developed to compensate cases when absence data are not available (see Phillips and Elith (2013) or Hastie and Fithian (2013) for further discussion).

Environmental data (explanatory/predictor variables)

There are three categories of environmental data which can be used in SDMs targeting marine species: (i) seafloor topographic variables, (ii) physical and chemical oceanographic variables, and (iii) biological variables. One of the main topographic variables used is bottom depth. Several variables can be calculated using bottom depth data such as slope, bottom complexity and isobath. Application of topographic variables to studies on mobile vertebrate was reviewed by Bouchet *et al.* (2014). Bottom substrate can also be considered as a topographic variable. Distance from topographic features, such as a coastline or a shelf break, can be used as explanatory variables, as these can be assumed to be potentially informative aliases or proxies for biological/ecological processes influencing the distribution of target species. Examples of physical and chemical oceanographic variables are temperature, salinity, sea surface height, current speed and direction, frontal boundary, mixed layer depth, dissolved chemicals and sea ice concentration; some of these variables were reviewed by Hobday and Hartong (2014). Chlorophyll-*a* (Chl-a) concentrations and prey densities are commonly used biological variables, but information on predators and competitors can also be used as variables if available. Some of the satellite-derived sea surface variables, such as temperate (SST), sea-surface height (SSH) and its anomalies (SSHa), current and sea ice concentrations, are available through a number of agencies without cost for non-commercial use (e.g., National Snow and Ice Data Center: www.nsidc.org). Satellite data are potentially useful explanatory variables for baleen whales as spatial coverage of most of them are global, often with frequent re-sampling over days, weeks or months. However, some of satellite data such as Chl-a and SST are susceptible to cloud cover (i.e. missing data due to cloud cover). To overcome such a problem, output from ocean models are now being used in SDM (e.g. Becker et al., 2016). Output of ocean models can be used both retrospective and forecast analyses. Geographic coordinates (e.g. longitude and latitude) might be used as explanatory variables to take account of spatial variations that cannot be captured by environmental factors alone, but, of course, such relationships are of little or no use in testing ecological hypotheses if there are no inferences upon the actual underlying ecological processes at work across geographical space. As well, spatial coordinates are critical for analyses that can account for spatial autocorrelation amongst environmental variables (e.g. Mantel's tests) when considering their effect on distribution patterns of species (e.g. Friedlaender *et al.*, 2006; see below for more details).

Spatial and temporal scales of data

It is ideal that both response and explanatory variables are recorded simultaneously. However, necessary environmental data might not be recorded at the time of observation of species because of availability of measurement instruments or logistical constraints. Furthermore, environmental data covering the entire target area is required to predict, via interpolation and/or extrapolation, the spatial distribution of the target species. In this way, environmental data are generally more useful than geographical coordinates for interpolation and/or extrapolation if one assumes relationships between environmental covariates and densities of target species remain stable over the study area. Principally, selection of spatial and temporal scales of environmental data should be based on sound ecological reasoning as well as objective of the analysis. However, availability of data could limit such a selection. For instance, if the coarsest spatial and temporal scales of environmental data are 5×5 km grid cell and month, respectively, resolutions of other data with finer resolutions should be resampled to match the coarsest resolutions.

Types of SDMs

Several types of statistical models are used as SDMs. These models can be classified into three methods according to Hijmans and Elith (2017): (i) regression methods, (ii) profile methods and (iii) machine learning methods. However, some methods are not easily slotted into this classification, and these are grouped under other methods in this manuscript. Ensembles of models are also mentioned here briefly. As mentioned in the previous section, SDMs can be classified based on type of response variables. Classification of some major SDMs by statistical models and types of response variables is summarized in Table 1. The presence/pseudo-absence approach can be considered as a special case of the presence/absence approach. Statistical models which can deal with presence/absence data can basically deal with presence/pseudo-absence data.

Regression methods

A general linear model is a fundamental regression method that assumes a linear relationship between response and explanatory variables. A general linear model assumes the response variable is continuous and normally distributed. However, relationship between spatial distribution of species and their environment could be more complex. A generalized linear model (GLM; McCullagh and Nelder 1989) allows a relaxing of the assumptions of a general linear model. A GAM (Wood, 2006) further allows non-linear relationships. A GAM is now commonly used as SDMs for whales to deal with such complexities. GLM and GAM require presence/absence data. Abundance and density, instead of presence data, can also be used in these models. GAMs are particularly favoured owing to their flexibility for allowing relationships (existing or, admittedly, spurious) between response and independent variables to drive the fitting process, and not pre-conceived constraints on model format. As noted above, a GAM-based spatial model for estimating whale abundance (density surface model; DSM) has also been developed (Miller et al., 2013).

Profile methods

In the profile method, distinctive environmental conditions at locations of presence are identified through values, such as means and ranges. The profile method is best suited to presence-only data. The habitat suitability index (HSI) (Verner *et al.*, 1986), the bioclimatic analysis and prediction system (BIOCLIM) (Busby, 1991), and DOMAIN (Carpenter *et al.*, 1993) are used as SDMs. Relative environmental suitability model (RES) (Kaschner *et al.*, 2006) can be considered as a type of HSI.

Machine learning methods

In the machine learning method, species distributions in relation to their environment are determined based on certain rules. Basically, presence/absence data are required in the machine learning methods. Decision-tree based models such as random forest (RF) (Breiman, 2001) and boosted regression tree (BRT) (Elith *et al.*, 2008) are categorized as machine learning methods (see also De'ath (2007) and De'ath and Fabricius (2000) for further details). Other models, such as Genetic Algorithms (GA) (Holland, 1975), Bayesian Networks (BN) and support vector machine (SVM) (Guo *et al.*, 2005) are also used as SDMs. A maximum entropy-based method called MaxEnt (Elith *et al.*, 2011) is specifically designed to use presence/background data. Because models categorized as the machine learning method are diverse, their characteristics are briefly described below.

Random forests

Random forests (RF) is part of a family of robust methods known as non-parametric. RF was developed by Brieman (2001) and, like other machine learning techniques, it has quickly become popular among the data science community because of its ability to model the complex structure of high-dimensional data sets. At its core, RF is a classification technique that combines many single decision trees in an embedded way to calculate the importance of each predictor. RF is also considered an ensemble method because it aggregates the results of multiple, independently generated classification trees into an averaged prediction.

RF is a classification and regression tree (CART) method based on bagging. Bagging generates n bootstrap samples, builds a model for each, and then averages the resulting models across bootstrap aggregates. The RF algorithm is executed by bootstrapping (with replacement) 63% of the data and generating a "weak learner" based on a CART for each bootstrap replicate. Within the pre-set specification (e.g. node depth and number of samples per node) each CART is unconstrained (grown to fullest) and prediction is accomplished by taking the "majority votes" across all nodes in all random trees. At each replicate the data not used to construct the tree [out of bag (OOB)] are used for validation, providing a quasi-independent validation of model fit. Covariates are randomly selected at each node and variable importance is assessed using the mean decrease in accuracy (MDA) by dividing the standard error by the misclassification rate. The number of covariates randomly selected at each node is defined by m (commonly defined as the square root of the number of covariates). The contribution of covariates can also be obtained with the Gini Index. Each time a node split occurs based on a particular variable, the Gini impurity criterion for the two descendent nodes is less than the parent node. The Gini index is calculated by summing the Gini decreases for each individual variable over all trees in the forest.

Owing to its unique approach to modelling, the use of RF in SDM has proven robust and stable. For marine taxa, several studies have successfully applied RF to seabird survey (Oppel and Huettmann 2010, Oppel et al. 2012, Renner et al. 2013, Liske et al. 2014) as well as tracking data (Scales et al. 2015). These studies should constitute excellent background material for applications with cetaceans. An area of focus in SDM has been the assessment of the performance of RF relative to various other modelling techniques (Marmion et al. 2009a, b). RF also has been used to assess the relative performance of models trained on abundance data and those trained on presence-absence data (Howard et al. 2014). RF is included in the suite of techniques for ensemble forecasting of species distributions [along with Generalized Additive Models (GAM), Maximum Entropy (MaxEnt), and Boosted Regression Trees (BRT)] that are implemented in the extremely popular BIOMOD platform (Araujo and New 2007, Thuiller et al. 2009, Thuiller 2014). To the best of our knowledge, RF has not been used directly as a SDM approach with cetaceans, but this is only a matter of time since the methodology is well established and is well suited for cetacean data sets, either as abundance or as presence-absence. Because seabirds have great similarities with cetaceans in terms of ecology and data collection techniques, the studies of Oppel and Huettmann (2010), Oppel et al. (2012), Renner et al. (2013), Liske et al. (2014), and Scales et al. (2015) provide excellent background material for applications of RF as SDM with cetaceans. We also note that a recent SDM study used GAM to generate habitat-based cetacean density predictions for a large number of cetacean species in waters of the U.S. Atlantic and Gulf of Mexico, and then implemented RF to resolve ambiguity in models containing similar species due to difficulties in field identification (Roberts et al. 2016).

Genetic algorithms

Originally developed by Holland (1975), genetic algorithms (GA) are a rule-based optimisation technique for supervised classification. They are thought to work well in situations involving complex interactions between many variables, particularly when derivative-based techniques prove problematic and complete enumeration of a search space for an optimal solution is not practical (Haupt and Haupt, 2004). Advances in computing power over the last 30 years have seen widespread uptake of GA in a variety of disciplines, including for species distribution modelling (SDM) in ecology. Stockwell and Noble (1992) introduced the earliest system for GA-based rule-set classification of species distributions according to environmental predictors. Termed GARP (Genetic Algorithm for Rule set Production), the Stockwell framework has proved a popular modelling choice in the applied ecological literature, showing over 1,000 Google Scholar citations over the period 2000-2015 for Stockwell and Peters' (1999) exposition of the technique. Variants of GARP, DK-GARP (Stockwell and Noble, 1992) and OM-GARP (Muñoz et al., 2009) were also developed

Genetic algorithms are a stochastic search algorithm motivated by the success of biological evolution and the processes of natural selection. They work to iteratively "evolve" a set of rules, evaluated against an objective function, to achieve some defined optimisation goal (Haupt and Haupt, 2004). For purposes of clarity, it is useful to immediately frame this in the context of SDM, in which:

> i) The optimisation goal is to obtain the best predictive model of species occurrence based on point locations of species observations and a suite of geographically referenced environmental covariates;
>
> ii) The objective function is typically formulated as the predictive accuracy of a model evaluated against an independent set of test data (i.e. data not used in forming the model rules); and,
>
> iii) The set of rules defines the way predictors relate to the probability of species' occurrence.

At each iteration (i.e. generation), selection is achieved by choosing optimal rule sets from among many candidates based on improvement of the objective function. To a restricted set of the best solutions, GA then applies analogues of reproductive processes, such as mutation and crossover, to create subsequent generations of solutions (Mitchell, 1996). Mutation modifies one or more aspect of a rule from its initial state in order to decrease the likelihood that candidate solutions become trapped at local minima. Genetic recombination is simulated via crossover operations, in which characteristics from two or more "parent" solutions are combined to form a "child" for the next generation. The basic premise of crossover is that combining several good solutions has the potential to come up with an even better one. Candidate solutions evolve through iterative modification of rule sets, either for a pre-specified (large) number of iterations or until no further improvements in the objective function can be found. GARP belongs to the class of SDM known as presence/background methods (sometimes called presence/pseudo-absence; see Renner *et al*., 2015 for a review of terminology) that make use of species occurrence data but must infer species' absences from areas within the study domain where the species has not been reported (Franklin, 2009).

Cetacean SDM based on GA are relatively rare, but not unknown. For a variety of fish and cetacean species, Ready *et al*. (2010) compare several different methods for SDM including GARP, RES, MaxEnt, GLM and GAM. Of direct relevance to this review, data were modelled for southern bottlenose whale (*Hyperoodon planifrons*) and fin whale (*Balaenoptera physalus*) from the Southern Ocean, and harbour porpoise (*Phocoena phocoena*) from the North East Atlantic. Results from different SDM methods were variable across fish and mammal species, but GARP consistently performed poorly. MacLeod *et al*. (2008) compared four SDM methods for modelling the occurrence of the harbour porpoise (*Phoceoena phocoena*) in the Sea of Hebrides, Scotland. The methods compared were GARP (using the default settings of DK-GARP), GLM, a PCA-based approach (Robertson *et al*., 2001) and ENFA (see below for ENFA). The study showed that all four techniques produced statistically equivalent results, with point estimates of AUC in the range 0.74-0.82. Similarly, spatial predictions for the 12 sub-areas were strongly and significantly correlated between all four modelling techniques.

There are no known advantages to using GA (including GARP) in relation to SDM studies of cetacean species. Many of the issues associated with applying SDM to cetaceans are unlikely to be able to be directly addressed through a GA framework, including issues related to paucity of data, observer biases, and a lack of direct links between sightings and environmental correlates during migratory behaviour. In light of these limitations, including the poor predictive performance of GARP shown in several studies, the approach is currently not recommended for developing SDM for cetacean species.

Support Vector Machines

The Support Vector Machine (SVM) method is a type of machine learning method for statistical pattern recognition. That is, supervised learning is performed when a training dataset is analyzed to develop an algorithm that is used to assign results to new examples in a test dataset. SVMs were originally introduced as a binary classifier (Vapnik 1998). Since then it has been extended to situations involving multiple classes, 1-class present only (e.g., untrained algorithms), partially identified classes, and even regressions. Basically SVM uses a functional relationship known as a kernel to map data onto a new hyperspace in which complicated patterns can be more simply represented. Because SVM are not based on characteristics of statistical distributions there is no theoretical requirement for observed data to be independent, thereby overcoming the problem of auto-correlated observations, although model performance will be affected by how well the observed data represent the range of environmental variables. In its classical implementation, a 2-class SVM uses two classes (e.g., presence/absence) of training samples within a multidimensional feature space to fit an optimal separating hyperplane in each dimension. In this way, SVM tries to maximize the margin that is the distance between the closest training samples, or support vectors, and the hyperplane itself. The classification can be modeled with a linear or non-linear algorithm. For example, presence of known locations of rare tree species and absent locations without these rare trees, along with the physical and biological characteristics of both types of locations, were used to predict the potential spatial distribution of the rare tree species (Pouteau et al. 2012). Distribution maps of a fish species were modeled from presence/absence data and 19 physical-chemical and environmental variables from freshwater rivers in northern Italy (Tirelli et al. 2012). The 2-class SVM has been generalized to a multiclass SVM to accommodate data that have been labeled into a finite set of classes. The dominant approach for doing so is to reduce the single multiclass problem into multiple binary classification problems, though one step likelihoods have also been attempted. Typically if absence data are not available or unreliable, then pseudo-absence data are generated. An example of presence only data are museum-collected locations of animals. An example of potentially unreliable absence data is absence of a mobile species since it is possible the survey just by chance did not see a mobile animals in a particular type of habitat or absence of an invasive species that has not yet spread to an area. To analyze the present-only format data Scholkopf et al. (1999) developed a one-class SVM. For example, Guo et al. (2005) used the one-class SVM methods to map the potential distribution in California of a tree virulent pathogen called Sudden Oak Death. Drake et al. (2006) used presence of 106 species in mountains of the Swiss alps along with nine environmental variables to model their distributions, thus interpreting this as the species ecological niche (a multidimensional environmental space). If the data are not labeled into categories or only some of the data are labeled, the SVM methodology was expanded to support vector clustering (SVC) which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. An advantage of this method is there are no assumptions on the number or shape of the clusters in the data (Ben-Hur et al. 2001). In SVC, data points are mapped from data space to a high dimensional feature space using a kernel function. In feature space the smallest sphere is searched for that encloses the image of the data using the Support Vector Domain Description algorithm. This sphere, when mapped back to data space, forms a set of contours which enclose the data points. These contours are interpreted as cluster boundaries, and points enclosed by each contour are associated by SVC to the same cluster. The basic idea behind support vector regression is to map the data into a high-dimensional feature space via a nonlinear mapping and do linear regression in this space. In essence, linear regression in a high dimensional feature space corresponds to nonlinear regression in a low dimensional space.

SVMs have been applied successfully to text categorization, handwriting recognition, gene-function prediction, and remote sensing classification, demonstrating the utility of the method across disciplines, proving that SVMs produce very competitive results with the best available classification methods. However, they have been applied to ecological predications only in the last decade and not frequently (examples were mentioned above). To date, SVMs have not been used as SDMs of cetaceans. Though, it appears to be an appropriate tool to investigate developing cetacean SDMs for rarely encountered species or in situations with limited or unreliable effort information.

Bayesian Networks

Bayesian networks (BN) are a kind of probabilistic graphical models that correlative and causal relationship among variables are represented graphically and probabilistically. BN are categorized as a kind of machine learning methods. BN also called in different names like directed Acyclic Graphical Models, Bayesian belief networks and Bayes network. Text books of BNs are available such as Nielsen and Jensen (2009), Pourret *et al.* (2008) and Scutari and Denis (2014). Several reviews and guidelines for BNs in the context of environmental and ecological studies are also available (Aguilera *et al.*, 2011; Chen and Pollino, 2012; Marcot *et al.*, 2006; McCann *et al.*, 2006; Uusitalo, 2007). BN have used as species distribution models (SDMs) since early 2000s. BN mainly consist of qualitative and qualitative components. In the qualitative component, causal relationships among variables are represented as directed acyclic graphs (DAGs). In DAGs, nodes (variables in ellipses) are linked by arcs (also called as edges and arrows) to show causal relationship between nodes. The initial structures of DAGs can be constructed based on known causal relationship (e.g. information from literature) and/or expert knowledge. In the qualitative component, degree of belief expressed as probability of a node in a particular state given states of parent node assuming that conditionally independent of all its non-descendants, given its parents.

BN have been applied as SDMs to a variety of inland vertebrates since early 2000s (e.g. Raphael et al., 2001; Tantipisanuh et al., 2014). The response variables of these studies were not abundance but presence and absence. Number of published papers using BN as SDMs is small in comparison with other machine learning methods. To date, BN have not been used as SDMs of cetaceans. Because of the limitation that variables should be discretized in some extent, utility of BN for management of cetaceans could be limited as detailed information is lost due the discretization. However, BNs could be useful tool for exploratory research to investigate causal relationship among variables based on expert knowledge which cannot be handled by other SDMs.

Maxent

Maxent is, at its most basic level, a method for making predictions or inferences from incomplete information (Phillips et al. 2006). Maxent generates presence-only models of species distributions by estimating the probability of distribution relative to maximum entropy (i.e. uniformity). The probability of a species occurrence is constrained as a function of environmental variables included as predictor variables. More precise and detailed explanations of the mathematical models used in Maxent are reviewed in Phillips et al. (2006), Phillips and Dudik (2008), and in Elith et al. (2011). In order to generate a model of a species' environmental requirements, Maxent uses a set of occurrence localities (presences). The environmental features that can be used in Maxent to predict a species distribution can be derived from both continuous and categorical variables. Maxent employs a number of features to fit a function of the covariates that include linear, product, quadratic, hinge threshold, and categorical. Explanations of the differences in how these are used by Maxent to derive relationships to covariates are provided in Elith et al. (2011). To date, Maxent has been used in a number of whale studies as reviewed in this paper. Because of the types of data that are required (or not required), it is amenable to a wide range of objectives, study areas, and species.

Other methods

Hierarchical Bayesian model (HBM) is used as a SDM to deal explicitly with heterogeneity in detection of species (Royle and Dorazio, 2008). Other advantages of HBM are as follows: allowing for the explicit propagation of uncertainty, and for several sub-models to be seamlessly integrated; allowing for the estimation of abundance (through Distance sampling) simultaneously with the estimation of association with environmental variables as separate sub-models (see Pardo *et al.*, 2015 for more details).

In Ecological Niche Factor Analysis (ENFA), species distribution is characterized to take account of contrasts between environmental conditions at presence of species and the background in (Hirzel *et al.*, 2002).

Ensemble modelling

It was documented that different SDMs applied to the same data sets created different prediction results (Elith *et al.*, 2006; Segurado and Araújo, 2004). Ensemble modelling (Araújo and New, 2007) is applied to deal with errors and uncertainties of each SDM. Four ensemble methods are proposed by Araújo and New (2007): (i) bounding box or generating a consensus forecast for small ensemble size; (ii) showing number of models forecasting presence using histogram; (iii) showing probability density function of likelihood of species presence for large ensembles and (iv) measuring central tendency (e.g. mean and median).

Notes on input data

Sampling of response variable

It is desirable that the response variable (occurrence of species) be randomly sampled from a target area based on robust statistical design, or, if this is too strong an assumption, at least randomly sampled in relation to the environmental covariates. Even if data are not sampled randomly, it is ideal to ensure broad coverage, perhaps via a systematic survey design, throughout a target area, particularly if distributions of a target species have, in that area, have previously not been studied. Obviously, the survey costs, logistics, and requirements for minimum useful sample sizes will also influence survey design. However, commercial catch and telemetry data could violate such assumptions as the nature of these data sets is basically non-random, and spatial and temporal coverage is likely to be limited. It is necessary to check whether the data are potentially biased, both spatially and temporally, before the modelling. Careful interpretations of the outcomes of SDMs are necessary if there are such biases. It is desirable to use both presence and absence data in SDMs whenever these data are available as it is expected that these data contained much information on environmental conditions that constrain presence of species in comparison with presence only data.

Reliability of response variable

Occurrence of species must be recorded correctly. However, it might not be in some cases such as miss-identification of species (Conn *et al.*, 2013) and imperfect detection (Laake *et al.*, 2008). Appropriate analytical treatment is necessary if either are suspected. Much has been written about biases introduced via heterogeneity in detection probabilities in fauna and flora surveys (e.g., Borchers *et al.* (2006), Ramsey and Harrison (2004) and Thomson et al. (2012) to name but a few examples).

Spatial autocorrelation of response variable

Spatial autocorrelation (SAC) in the context of SDMs was reviewed by Dormann *et al.* (2007). SAC occurs when data sampled in close proximity are not independent from each other. In a general sense, samples that are too similar, in this example, because they are adjacent in space and time, will yield falsely low variance estimates. Independence among data, which is assumed in standard statistical models, is violated if SAC exists and it can lead to type I error. Existence of SAC can be checked by some indices such as Moran's I and Geary's C, or explored using geostatistical or mixed-effects modelling. Several statistical models which can deal with SAC are available (Table 2). Whilst it is likely that telemetry data are susceptible to SAC, as the data obtained continuously from tagged species, it is possible to minimise this problem via methods such as blocking in space and time (Aarts et al. 2008).

Ecological validity of explanatory variables

Ideally, selection of explanatory variables should be at least broadly based on ecological reasoning, which can be obtained qualitatively as expert knowledge. However, the selection would be arbitrary if such information is not available before the modelling. In a predictive modelling context, there may be little or no desire to capture information about actual ecological/biological processes—but if it is captured in models, this can probably only help. With predictive modelling, the aim is produce models optimised for accurate predictions within the bounds of the system of interest (i.e., with little to no extrapolation). Ideally, predictor variables are cheap to collect, and plentiful, as the more information will lead to more accurate predictions—any relation they have to the variable of interest (here, the presence or densities of species) can be purely coincidental, just as long as it is useful for accurate prediction. For more thorough discussions on the differences between predictive and explanatory modelling, see Mac Nally (2000) and Shmueli (2010).

Collinearity among explanatory variables

Collinearity in the context of SDMs was reviewed by Dormann *et al.* (2012). Collinearity implies that some of explanatory variables, especially in regression methods, are related. If collinearly related variables exist in a model, explanatory power of one of the collinearly related variables might be reduced and/or the model can be unstable. Collinearity can be serious problem when a selected model is used for prediction where structure of collinearity is unknown. Existence of collinearity can be checked using some indices such as variance inflation factor (VIF). Several methods which can be applied before modelling or during modelling are available to deal with collinearity, as reviewed in Dormann *et al.* (2012). Hierarchical partitioning is one method that assists in teasing out the effects of collinearity with the aim of identifying potentially important explanatory variables (Mac Nally 2000).

Model Evaluation

Model evaluation methods for presence/absence data was reviewed by Fielding and Bell (1997). Area under the curve (AUC) of receiver operating characteristic curve (ROC) is one commonly used method. For abundance data, predictions and overall model performance were compared using explained deviance, average squared prediction error (ASPE), and ratios of observed to predicted densities to identify the best models (Forney *et al.*, 2012).

REVIEW OF APPLICATIONS OF SDMS TO BALEEN WHALES

A total of 48 published in scientific journals from December 1997 to December 2016 are considered to review applications of SDMs to baleen whales.

Target species, regions and areas

Target species, regions and areas considered in the published papers are summarized in Table 3. SDMs were applied to all baleen whales except pygmy right (*Caperea marginata*) and Omura's (*B. omurai*) whales. SDMs were applied to a wide variety of regions and areas, but there are few applications in the Indian Ocean and South Pacific. No study was conducted in the South Atlantic. Some papers dealt with multi-species and/or multi-regions.

Data

Response variable

Response variables used in the published papers are summarized in Table 4. Data sets obtained by dedicated sighting surveys were used in 24 papers followed by data sets obtained by opportunistic sighting surveys (18 papers). In this review, surveys conducted based on the DISTANCE sampling method were defined as dedicated surveys. Catch data and published data were also used in SDMs. Count (number of animals or abundance) data sets were used in 26 papers followed by presence only (12 papers) and presence/absence (9 papers) data sets.

Explanatory variables

Explanatory variables used in the published papers are summarized in Table 5. A total of 31 types of explanatory variables were used in the published papers. Bottom depth was the most commonly used variable in the SDMs (39 papers) followed by sea surface temperature (SST) (31 papers), seafloor slope (24 papers), surface chlorophyll-*a* concentrations/primary production (23 papers), longitude (17 papers), latitude (16 papers), distance to shore (12 papers) and sea surface height (12 papers). Other variables were used in less than 10 papers.

Statistical models used as SDMs

Statistical models used in the published papers are summarized in Table 6. Only two papers considered two different statistical models. Other papers used only one statistical model per analysis. GAMs, including one paper that used mixed-effects GAM (GAMM), were the most used models (27 papers) followed by MaxEnt (10 papers). Other models (BRT, CART, ENFA, GLM, HBM, logistic regression and RES) were used in 4 or fewer papers. One paper used a geostatistical method, kernel density smoothing, to predict spatial distribution of baleen whales (Laidre *et al.*, 2010). Forney *et al.* (2012) and Víkingsson *et al.* (2015) used GAMs as SDMs, but the spatial distributions were predicted by geostatistical methods, namely inverse distance weighting (IDW) and kriging, respectively. SACs were considered in 6 papers, using various methods. Collinearity was considered in 7 papers. Resolutions of grid cells used for spatial distributions were varied among papers (approximately from 1 to 50 km).

Parameter setting for GAMs

Some key parameter settings for GAMs are reviewed here as GAMs have been used frequently as SDMs for baleen whales. Common parameter settings for GAMs were summarized in Table 7. Some of the settings are specific to the mgcv package (Wood, 2006) of the R software (R Development Core Team, 2015).

Response variables based on animal count data often suffer from overdispersion and zero inflation (i.e., more variation than can be explained by a Poisson distribution). To account for this, GAMs (and GLMs) can use quasi-Poisson, negative binomial or Tweedie distributions.

Some GAM parameter settings were not reported in several papers but it was likely that they used default settings (which, in general, should not lead to completely incorrect outcomes as Simon Wood, the author of mgcv, worked hard to set the defaults to represent an overall reasonable compromise across manifold complex considerations). Model selections were based on either GCV, AIC, BIC (Bayesian information criterion) or REML (restricted maximum likelihood).

Whilst GAMs allow flexibility, and for features of the data to drive model parameterisation, it is often the case that relationships between explanatory and response variables can be either under or over-described or ‘smoothed’. For example, a GAM might indicate that density of some species has three or more ‘maxima’ over a range of a given environmental covariate. If the covariate represents a biological gradient, is it likely that a particular species has several optimal values over such a variable? Therefore, care does need to be taken in pre-defining (or not) assumptions about ‘wiggliness’ of densities along gradients.

A GUIDELINE FOR SPECIES DISTRIBUTION MODELLING

Although there is a number of guidelines for modelling in the context of environmental/ecological studies, ten iterative steps in development and evaluation of models proposed by Jakeman *et al.* (2006) can be considered as a good starting point for development of a guideline for SDMs applied to cetaceans. Written statement of these steps will help reviewers. The following are general comments on each steps for the purpose of the review of SDMs.

First step: Definition of purposes for modelling

In border context, there are at least three purposes of development of SDMs applied to cetaceans: (1) estimation of spatial abundance (2) estimation of spatial distribution and (3) investigation on ecological questions (e.g. habitat requirement). These purposes might involve interpolation (estimation within a target [survey] area) and extrapolation (prediction outside of a target area and future projection). However, these two words can be defined differently: interpretation can be regarded as estimation within line transect strip (length of transect times effective search width) while extrapolation can be regarded as estimation within a target area. These three are not mutually exclusive and some of statistical models can address these at once. Nevertheless, distinction of main purposes for modelling is important because they affect details of subsequent model development and evaluation steps. Former two purposes are more related to in-depth assessment and management of stocks while the third purpose is more related to ecological questions.

Second step: Specification of the modelling context

According to the Jakeman *et al.* (2006), following 9 points should be considered at this step: (1) the specific questions and issues that the model is to address, (2) the interest group, including the clients or end-users of the model, (3) the outputs required, (4) the forcing variables (drivers), (5) the accuracy expected or hoped for, (6) temporal and spatial scope, scales and resolution, (7) the time frame to complete the model as fixed, (8) the effort and resources available for modelling and operating the model and (9) flexibility. Jakeman *et al.* (2006) considered that the crucial point at this step is (6).

Specific questions and issues that the model is to address
Following are some examples of specific questions and issues that the SDMs to address:

- Spatial abundance estimation for the purpose of RMP
- Investigation on reasons of change in spatial abundance and distribution for the purpose of IA
- Identification of distribution area of whales to reduce ship strikes in a certain area
- Investigation on habitat requirements for ecological study

Interest group, including the clients or end-users of the model
In general, interest groups of SDMs could consist of managers, scientists, fishers, conservation groups and general public though specific combination would be varied from case to case.

Outputs required
Primary outputs from SDMs are estimated maps of probability of occurrence and/or abundance. Point estimate can also be obtained in the case of abundance. Importance (rank) of environmental variables affecting occurrence/abundance can also be obtained from models.

Forcing variables (drivers)
This is not applicable to SDMs as forcing variables are not used in the models in general.

Accuracy expected or hoped for
Acceptable level of accuracy should be determined by discussion before conducting modelling but it might be changes on the course of analysis.

Temporal and spatial scope, scales and resolution
Specification of temporal and spatial scope, scales and resolution is closely tied with the purpose of modelling. For example, if one aims to estimate spatial abundance and distribution of a stock in a particular season, broader spatial area should be covered by the modelling. For instance, in the case of the third circumpolar survey under IWC SOWER, it took approximately 40 days by two vessels to cover an area in a 30° longitude sector from ice edge to 60°S. In such a case, temporal scale of environmental data (e.g., temperature) for modelling might be restricted to month or seasonal mean data. In contrast, if one aims to estimate spatial abundance and distribution in a local area (e.g., a bay), may be covered in a few days. In that case, environmental data with high temporal resolution might be used but only a fraction of a stock might be studied. Specification might be limited by available environmental data. For example, sea surface temperature derived from satellite data is commonly used as an environmental data in modelling. However, both temporal (e.g. observed period) and spatial (e.g. grid size and cloud cover) coverages are limited.

Time frame to complete the model as fixed, for example, by when it must be ready to help a decision
Time frame should be determined by discussion before conducting modelling but it might be changed on the course of analysis.

Effort and resources available for modelling and operating the model
Identification of effort and resources available for modelling is important to set time frame and required budget. Consideration of operation of the model might be necessary if constructed models are applied to new environmental data (e.g. temperature) continuously.

Flexibility; for example, can the model be quickly reconfigured to explore a new scenario proposed by a management group
Flexibility of models used as SDMs should be described although most of them are reasonably flexible for reconfiguration.

Third step: Conceptualisation of the system, specification of data and other prior knowledge
A reasonable hypothesis about the relationship between explanatory variables (usually environmental variables) and response variable (presence/absence or abundance) should be provided. It is directly related to selection of explanatory variables for an initial model.

Response variable are one from the followings:

- Presence/absence: Typically collected by sighting survey (either dedicated or opportunistic) as sighting effort data are required.
- Presence only: Typically collected by satellite tags as the data provide only location of cetaceans.
- Abundance: Typically collected by dedicated sighting survey which records distance and angle of sightings, and school size to calculate effective half width and mean school size.

Type of response variable has a strong influence on selection of model features and families.

A variety of explanatory variables may have been used in a given study. The details of data should be provided. The followings are some of the examples:

- *In-situ* environmental data: Environmental data recorded during field surveys, such as water temperature obtained by CTD and prey density obtained by echosounder are used in SDMs. Interpolation and/or extrapolation of data for a target area are necessary as these data are recorded along track lines in most cases.
- Satellite data: Environmental data obtained satellite are commonly used in SDMs as the data have wide coverage both temporally and spatially. Types of data include such as SST, SSH, sea surface chlorophyll-a concentrations (chl-a) and sea ice concentrations. Interpolation and/or extrapolation of SST and chl-a data might be necessary in cases of missing values due to cloud cover. Secondary data products such as thermal fronts calculated using satellite data are also available for some regions.
- Terrain data: Digital bottom depth data and variables calculated using the data (e.g. slope) are used in SDMs. Distance from terrain futures such as coastline are also used.
- Ocean model data: Output from ocean model data (e.g. Regional Ocean Modeling System [ROMS]) are used in SDMs.
- Climatological data: Climatological data (e.g. World Ocean Atlas published by NOAA) are used in SDMs.

At this stage, considerations on spatial autocorrelation of response variable and collinearity among explanatory variables are also required especially for regression models.

Forth step: Selection of model features and families
Although a number of statistical models can be used as SDMs, selection of families (i.e. specific statistical models) is limited by features (e.g. types of variables and linear/nonlinear functions). Description of reasons why a particular model is selected is desirable. It is preferable to use several models and compare the results. An alternative choice could be ensemble modelling if the primary objective is estimation of spatial abundance and distribution. However, a major drawback of ensemble modelling is that it cannot be utilized for ecological inferences.

Fifth step: Choice of how model structure and parameter values are to be found
Choice of model structure (i.e. relation between variables) can be inferred from prior scientific knowledge. However, the choice could be limited by availability of explanatory variables for SDMs. Methods to estimate parameter values are specific to each statistical model.

Sixth step: Choice of estimation performance criteria and technique
Each statistical model has unique methods for parameter estimation performance criteria and technique, and it should be described fully.

Seventh step: Identification of model structure and parameters
In many cases, this step just consists of dropping or adding of particular parameters to reduce or increase model complexity based on steps 5 and 6 above.

Eighth step: Conditional verification including diagnostic checking

There are generally two forms of verifications: quantitative and qualitative verifications. Qualitative (conceptual) verification is verification between a real system and a conceptual model, based on qualitative information such as expert knowledge. Quantitative (model) verification is verification between conceptual and quantitative models, based quantitative criteria such as goodness-of-fit and tests on various residuals.

Ninth step: Quantification of uncertainty

The uncertainty associated with abundance can be estimated via methods such as bootstrapping in GAMs. However, uncertainty associated with probability of occurrence has rarely been explored.

Tenth step: Model evaluation or testing (other models, algorithms, comparisons with alternatives)

Model evaluation using test data have been conducted for probability of occurrence based on AUC. However, it has rarely conducted for abundance. Comparison of results among different statistical models is recommended to evaluate them. Point estimate (e.g. abundance) comparison is relatively easy but ecological inference might be difficult if different models show different results (e.g. shape of response form).

DISCUSSION

A wide variety of statistical models as SDMs have been applied to most of the baleen whale species known worldwide. The results significantly contribute to expanding our knowledge of whale ecology and behaviour. It seems that applications of SDMs to baleen whales up to now are exploratory as most cetacean scientists are still in the capacity building stage in this field. In addition, the applications are limited where appropriate data sets for the modelling are available. Most studies only applied a statistical model, and comparison of results from applying different models to a data set has not been conducted fully. As mentioned above, it was reported that SDMs applied to same data sets created different prediction results (Elith *et al.*, 2006; Segurado and Araújo, 2004). Therefore, it is recommended that comparison of results of different models should be conducted in future studies. Furthermore, ensemble modelling should also be attempted to deal with errors and uncertainties of each SDM. However, it should be noted these studies only considered models to deal with probability of occurrence. A study was conducted to compare results among SDMs which use count data of seabird as a response variable (Renner *et al.*, 2013). Renner *et al.* (2013) also attempted ensemble modelling using count data. Oppel *et al.* (2012) also compared and attempted ensemble modelling using seabird data. We recommend such attempts should also be conducted using baleen whale data.

Construction methods of SDMs in the published literature, as applied to baleen whales, were not consistent. For instance, SACs and collinearity were not considered in most of papers, or, at least, not directly reported in such studies. It is recommended to consider at least these two factors when constructing SDMs for baleen whales as they might have significant effect on the results of modelling, dependent on model objectives.

GAMs were the most commonly used SDMs applied to baleen whales, probably because an established method in the cetacean literature (i.e., Forney 2000; Hedley *et al.*, 1999; Palka 1995) has been available since the 1990s. There are a number of settings for GAM modelling but there is no consensus method or documentation to help set them; although Forney *et al.* (2012) provides a good summary. The setting will be different for different data sets, but some guidelines are required to narrow down the choices. The R package, *dsm*, is now available for density surface modelling (Miller et al., 2017) and it allows testing of a range of settings. It is recommended that this point should be considered in on-going work of the RMP sub-Committee, with particular reference to development of a guideline for model based abundance estimation method. Setting for other models such as MaxEnt should also be considered in the future work.

REFERENCES

Aarts, G., MacKenzie, M., McConnell, B., Fedak, M. and Matthiopoulos, J. 2008. Estimating space‐use and habitat preference from wildlife telemetry data. *Ecography* 31: 140‐160.

Aguilera, P.A., Fernández, A., Fernández, R., Rumí, R. and Salmerón, A. 2011. Bayesian networks in environmental modelling. *Environ. Modell. Softw.* 26: 1376-88.

Ainley, D., Jongsomjit, D., Ballard, G., Thiele, D., Fraser, W., Tynan, C. 2012. Modeling the relationship of Antarctic minke whales to major ocean boundaries. *Polar Biol.* 35: 281-90.

Araújo, M.B., New, M. 2007. Ensemble forecasting of species distributions. *Trends Ecol. Evol.* 22: 42-7.

Ballard, G., Jongsomjit, D., Veloz, S.D., Ainley, D.G. 2012. Coexistence of mesopredators in an intact polar ocean ecosystem: The basis for defining a Ross Sea marine protected area. *Biol. Conserv* 156: 72-82.

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. *Endanger. Species Res.* 16: 97-112.

Becker, E., Forney, K., Fiedler, P., Barlow, J., Chivers, S., Edwards, C., Moore, A. and Redfern, J. 2016. Moving towards dynamic ocean management: How well do Modeled ocean products predict species distributions? *Rem. Sens.* 8: 149.

Beekmans, B.W.P.M., Forcada, J.F., Murphy, E.J., de Baar, H.J.W., Bathmann, U.V., Fleming, A.H. 2010. Generalised additive models to investigate environmental drivers of Antarctic minke whale (*Balaenoptera bonaerensis*) spatial density in austral summer. *J. Cetacean Res. Manage.* 11: 115-29.

Ben-Hur, A., Horn, D., Siegelmann, H.T. and Vapnik, V.. 2001. Support vector clustering. *J Mach. Learn. Res.* 2:125-37.

Best, B.D., Fox, C.H., Williams, R., Halpin, P.N. and Paquet, P.C. 2015. Updated marine mammal distribution and abundance estimates in British Columbia. *J. Cetacean Res. Manage.* 15: 9-26.

Best, B.D., Halpin, P.N., Read, A.J., Fujioka, E., Good, C.P., LaBrecque, E.A., Schick, R.S., .J., R.J., Hazen, L.J., Qian, S.S., L., P.D., P., G.L., McLellan, W.A. 2012. Online cetacean habitat modeling system for the US east coast and Gulf of Mexico. *Endanger. Species Res.* 18: 1-15.

Bombosch, A., Zitterbart, D.P., Van Opzeeland, I., Frickenhaus, S., Burkhardt, E., Wisz, M.S., Boebel, O. 2014. Predictive habitat modelling of humpback (*Megaptera novaeangliae*) and Antarctic minke (*Balaenoptera bonaerensis*) whales in the Southern Ocean as a planning tool for seismic surveys. *Deep Sea Res. I* 91: 101-14.

Borchers, D. L., Laake, J. L., Southwell, C. and Paxton, C. G. M. 2006. Accommodating unmodeled heterogeneity in double-observer distance sampling surveys. *Biometrics* 62: 372-78.

Bouchet, P.J., Meeuwig, J.J., Salgado Kent, C.P., Letessier, T.B., Jenner, C.K., 2014. Topographic determinants of mobile vertebrate predator hotspots: current knowledge and future directions. *Biol. Rev.* doi:10.1111/brv.12130

Breiman, L. 2001. Random Forests. *Machine Learning* 45: 5-32.

Busby, J.R. 1991. BIOCLIM - A Bioclimatic Analysis and Prediction System. pp.64-8 *In*: C.R. Margules, M.P. Austin (Eds.), *Nature Conservation: Cost Effective Biological Surveys and Data Analysis*. CSIRO, Canberra

Carpenter, G., Gillison, A.N., Winter, J. 1993. DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals. *Biodivers. Conserv.* 2: 667-80.

Chen, S.H. and Pollino, C.A. 2012. Good practice in Bayesian network modelling. *Environ. Modell. Softw.* 37: 134-45.

Conn, P. B., McClintock, B. T., Cameron, M. F., Johnson, D. S., Moreland, E. E. and Boveng, P. L. 2013. Accommodating species identification errors in transect surveys. *Ecology.* 94: 2607-18.

Corkeron, P.J., Minton, G., Collins, T., Findlay, K., Willson, A., Baldwin, R. 2011. Spatial models of sparse data to inform cetacean conservation planning: an example from Oman. *Endanger. Species Res.* 15: 39-52.

Cotté, C., Guinet, C., Taupier-Letage, I., Mate, B., Petiau, E. 2009. Scale-dependent habitat use by a large free-ranging predator, the Mediterranean fin whale. *Deep Sea Res. I* 56: 801-11.

Dalla Rosa, L., Ford, J.K.B., Trites, A.W. 2012. Distribution and relative abundance of humpback whales in relation to environmental variables in coastal British Columbia and adjacent waters. *Cont. Shelf Res.* 36: 89-104.

De'ath, G. and Fabricius, K. E. 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecolgy* 81: 3178-92.

De'ath, G. 2007. Boosted trees for ecological modeling and prediction. *Ecology* 88: 243-251.

De Kerckhove, D.T., Smokorowski, K.E., Randall, R.G. 2008. A Primer on Fish Habitat Models. *Can. Tech. Rep. Fish. Aquat. Sci.* 2817: 1-65

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S. 2012. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*: 27-46.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ralf, O., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M., Wilson, R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30: 609-28.

Drake, J.M., Randin, C. and Guisan, A. 2006. Modelling ecological niches with support vector machines. Journal of *Appl. Ecol.* 43: 424-432.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., E. Zimmermann, N. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129-51.

Elith, J., Leathwick, J.R., Hastie, T. 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77: 802-13.

Elith, J., Phillips, S.J., Hastie, T., Dudík, M., Chee, Y.E., Yates, C.J. 2011. A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions* 17: 43-57.

Fielding, A.H., Bell, J.F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* 24: 38-49.

Forney, K.A. 2000. Environmental models of cetacean abundance: reducing uncertainty in population trends. *Conserv. Biol.* 14: 1271-86.

Forney, K.A., Becker, E.A., Foley, D.G., Barlow, J. and Oleson, E.M. 2015. Habitat-based models of cetacean density and distribution in the central North Pacific. *Endanger. Species Res.* 27: 1-20.

Forney, K.A., Ferguson, M.C., Becker, E.A., Fiedler, P.C., Redfern, J.V., Barlow, J., Vilchis, I.L., Ballance, L.T. 2012. Habitat-based spatial models of cetacean density in the eastern Pacific Ocean. *Endanger. Species Res.* 16: 113-33.

Franklin, J. 2009. *Mapping species distributions: spatial inference and prediction*. Cambridge University Press, Cambridge.

Friedlaender, A.S., Halpin, P.N., Qian, S.S., Lawson, G.L., Wiebe, P.H., Thiele, D., Read, A.J. 2006. Whale distribution in relation to prey abundance and oceanographic processes in shelf waters of the Western Antarctic Peninsula. *Mar. Ecol. Prog. Ser.* 317: 297-310.

Friedlaender, A.S., Johnston, D.W., Fraser, W.R., Burns, J., Patrick N, H., Costa, D.P. 2011. Ecological niche modeling of sympatric krill predators around Marguerite Bay, Western Antarctic Peninsula. *Deep Sea Res. II* 58: 1729-40.

Gregr, E.J. 2011. Insights into North Pacific right whale Eubalaena japonica habitat from historic whaling records. *Endanger. Species Res.* 15: 223-39.

Gregr, E.J., Baumgartner, M.F., Laidre, K.L., Palacios, D.M. 2013. Marine mammal habitat models come of age: the emergence of ecological and management relevance. *Endanger. Species Res.* 22: 205-12.

Gregr, E.J., Trites, A.W. 2001. Predictions of critical habitat for five whale species in the waters of coastal British Columbia. *Can. J. Fish. Aquat. Sci.* 58: 1265-85.

Guisan, A., Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. *Ecol. Model.* 135: 147-86.

Guo, Q., Kelly, M., Graham, C.H. 2005. Support vector machines for predicting distribution of Sudden Oak Death in California. *Ecol. Model.* 182: 75-90.

Hamazaki, T. 2002. Spatiotemporal prediction models of cetacean habitats in the Mid-Western North Atlantic Ocean (from Cape Hatteras, North Carolina, U.S.A. to Nova Scotia, Canada). *Mar. Mamm. Sci.* 18: 920-39.

Hastie, T. and Fithian, W. 2013. Inference from presence-only data; the ongoing controversy. *Ecography* 36: 864-867.

Haupt, R.L. and Haupt, S.E. 2004. Practical Genetic Algorithms, 2nd Ed. John Wiley & Sons, New Jersey. 253p.

Hedley, S., Bravington, M. 2014. Comments on design-based and model-based abundance estimates for the RMP and other contexts. Paper SC/65b/RMP11presented to the 65b IWC Scientific Committee, May 2014. (unpublished). 9pp.

Hedley, S.L., Buckland, S.T., Borchers, D.L. 1999. Spatial modelling from line transect data. *J. Cetacean Res. Manage.* 1: 255-64.

Herr, H., Viquerat, S., Siegel, V., Kock, K.-H., Dorschel, B., Huneke, W.G.C., Bracher, A., Schröder, M. and Gutt, J. 2016. Horizontal niche partitioning of humpback and fin whales around the West Antarctic Peninsula: evidence from a concurrent whale and krill survey. *Polar Biol.* 39: 799-818.

Hijmans, R.J., Elith, J., 2017. Species distribution modeling with R (January 8, 2017), 78pp. Avairable from https://cran.r-project.org/web/packages/dismo/index.html (accessed on 25 April 2017)

Hirzel, A.H., Hausser, J., Chessel, D., Perrin, N. 2002. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology* 83: 2027-36.

Hobday, A.J., Hartog, J.R. 2014. Derived ocean features for dynamic ocean management. *Oceanography* 27: 134-4

Holland, J.H. 1975. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. University of Michigan Press. 183p.

Howard, C., Stephens, P.A., Pearce-Higgins, J.W., Gregory, R.D., Willis, S.G., 2014. Improving species distribution models: the value of data on abundance. Methods in *Ecol. Evol.* 5: 506–13.

IWC 2012. Requirements and guidelines for conducting surveys and analysing data within the revised management scheme. *J. Cetacean Res. Manage.* 13 (Suppl.): 509-17.

IWC 2015. Annex K1 Report of the Working Group on Ecosystem Modelling. *J. Cetacean Res. Manage.* 16 (suppl.): 277-90.

Jakeman, A.J., Letcher, R.A. and Norton, J.P. 2006. Ten iterative steps in development and evaluation of environmental models. *Environ. Modell. Softw.* 21: 602-14.

Kaschner, K., Watson, R., Trites, A.W., Pauly, D. 2006. Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. *Mar. Ecol. Prog. Ser.* 316: 285-310.

Keller, C.A., Garrison, L., Baumstark, R., Ward-Geiger, L.I., Hines, E. 2012. Application of a habitat model to define calving habitat of the North Atlantic right whale in the southeastern United States. *Endanger. Species Res.* 18: 73-87.

Kramer-Schadt, S., Niedballa, J., Pilgrim, J.D., Schröder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A.K., Augeri, D.M., Cheyne, S.M., Hearn, A.J., Ross, J., Macdonald, D.W., Mathai, J., Eaton, J., Marshall, A.J., Semiadi, G., Rustam, R., Bernard, H., Alfred, R., Samejima, H., Duckworth, J.W., Breitenmoser-Wuersten, C., Belant, J.L., Hofer, H., Wilting, A. 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. *Divers. Distrib.* 19: 1366-79.

Laake, J., Guenzel, R. J., Bengtson, J. L., Boveng, P., Cameron, M. and Hanson, M. B. 2008. Coping with variation in aerial survey protocol for line-transect sampling. *Wildlife Res.* 35: 289-98.

Laidre, K.L., Heide-Jorgensen, M.P., Heagerty, P., Cossio, A., Bergstrom, B., Simon, M. 2010. Spatial associations between large baleen whales and their prey in West Greenland. *Mar. Ecol. Prog. Ser.* 402: 269-84.

Laran, S., Gannier, A. 2008. Spatial and temporal prediction of fin whale distribution in the northwestern Mediterranean Sea. *ICES J. Mar. Sci.* 65: 1260-9.

Lieske, D.J., Fifield, D.A., Gjerdrum, C., 2014. Maps, models, and marine vulnerability: Assessing the community distribution of seabirds at-sea. *Biol. Conserv.* 172: 15–28.

McCann, R.K., Marcot, B.G. and Ellis, R. 2006. Bayesian belief networks: applications in ecology and natural resource management. *Canadian Journal of Forest Research* 36: 3053-62.

MacLeod, C.D., Mandleberg, L., Schweder, C., Bannon, S.M., and Pierce, G.J. 2008. A comparison of approaches for modelling the occurrence of marine animals. *Hydrobiologia* 612: 21–32.

Mac Nally, R. 2000. Regression and model-building in conservation biology, biogeography and ecology: The distinction between and reconciliation of 'predictive' and 'explanatory' models. *Biodivers. Conserv.* 9: 655-71.

Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L., Erickson, W.P. 2002. *Resource selection by animals: statistical design and analysis of field studies*, 2nd edition ed. Kluwer Academic Publishers, Dordrechut.

Marcot, B.G., Steventon, J.D., Sutherland, G.D. and McCann, R.K. 2006. Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Can. J. For. Res.* 36: 3063-74.

Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R.K., Thuiller, W., 2009a. Evaluation of consensus methods in predictive species distribution modelling. *Divers. Distrib.* 15: 59–69.

Marmion, M., Luoto, M., Heikkinen, R.K., Thuiller, W., 2009b. The performance of state-of-the-art modelling techniques depends on geographical distribution of species. *Ecol. Model.*220: 3512–3520.

McCullagh, P. and J. A. Nelder 1989. *Generalized Linear Models*. Oxford, Chapman and Hall.

Mitchell, M. 1996 An introduction to genetic algorithms. Massachusetts Institute of Technology. 158p.

Miller, D.L., Burt, M.L., Rexstad, E.A. and Thomas, L. 2013. Spatial models for distance sampling data: recent developments and future directions. Methods in Ecology and Evolution 4: 1001-10.

Miller, D. L., Rxstad, E., Burt, L., Bravingtion, M. V, and Hedley, S. 2017. Package 'dsm'. Available form https://cran.r-project.org/web/packages/dsm/index.html (accessed on 17 April 2017)

Monsarrat, S., Pennino, M.G., Smith, T.D., Reeves, R.R., Meynard, C.N., Kaplan, D.M. and Rodrigues, A.S.L. 2015. Historical summer distribution of the endangered North Atlantic right whale (*Eubalaena glacialis*): a hypothesis based on environmental preferences of a congeneric species. *Divers. Distrib.* 21: 925-37.

Monsarrat, S., Pennino, M.G., Smith, T.D., Reeves, R.R., Meynard, C.N., Kaplan, D.M. and Rodrigues, A.S.L. 2016. A spatially explicit estimate of the prewhaling abundance of the endangered North Atlantic right whale. *Conserv. Biol.* 30: 783-91.

Moses, E., Finn, J.T. 1997. Using Geographic Information Systems to predict North Atlantic right whale (*Eubalaena glacialis*) habitat. *J. North. Atl. Fish. Sci.* 22: 37-46.

Muñoz, M.E.S., Giovanni, R., Siqueira, M.F., Sutton, T., Brewer, P., Pereira, R.S., Canhos, D.A.L. and Canhos, V.P. 2009. "openModeller: a generic approach to species' potential distribution modelling". GeoInformatica. DOI: 10.1007/s10707-009-0090-7

Murase, H., Friedlaender, A., Kelly, N., Palacios, D.M. and Palka, D. 2015. A preliminary review of species distribution models (SDMs) applied to baleen whales. Paper SC/66a/EM3 presented to the 66a IWC Scientific Committee, May 2015. (unpublished). 18pp.

Murase, H., Friedlaender, A., Kelly, N., Kitakado, T., McKinlay, J., Palacios, D.M. and Palka, D. 2016. Progress report of the intersessional corresponding group 'Applications of species distribution models (SDMs)' since 66a IWC/SC. Paper SC/66b/EM4 presented to the 66b IWC Scientific Committee, June 2016. (unpublished). 27pp

Murase, H., Hakamada, T., Matsuoka, K., Nishiwaki, S., Inagake, D., Okazaki, M., Tojo, N., Kitakado, T. 2014. Distribution of sei whales (*Balaenoptera borealis*) in the subarctic–subtropical transition area of the western North Pacific in relation to oceanic fronts. *Deep Sea Res. II* 107: 22-8.

Murase, H., Kitakado, T., Hakamada, T., Matsuoka, K., Nishiwaki, S., Naganobu, M. 2013. Spatial distribution of Antarctic minke whales (*Balaenoptera bonaerensis*) in relation to spatial distributions of krill in the Ross Sea, Antarctica. *Fish. Oceanogr.* 22: 154-73.

Nielsen, T.D. and Jensen, F.V. 2009. *Bayesian Networks and Decision Graphs*. Springer New York. 448 pp.

Okamura, H., Matsuoka, K., Hakamada, T., Okazaki, M., Miyashita, T. 2001. Spatial and temporal structure of the western North Pacific minke whale distribution inferred from JARPN sightings data. *J. Cetacean Res. Manage.* 3: 193-200.

Oppel, S., Huettmann, F., 2010. Using a random forest model and public data to predict the distribution of prey for marine wildlife management. In: Cushman, S.A., Huettmann, F. (Eds.), Spatial Complexity, Informatics, and Wildlife Conservation. Springer, New York, pp. 151–163.

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. *Biol. Conserv.* 156: 94-104.

Palacios, D.M., Baumgartner, M.F., Laidre, K.L., Gregr, E.J. 2013. Beyond correlation: integrating environmentally and behaviourally mediated processes in models of marine mammal distributions. *Endanger. Species Res.* 22: 191-203.

Palka, D. 1995. Influences on spatial patterns of Gulf of Maine harbor porpoises. pp.69-75 In: L.W. Arnoldus Schytte Blix, U. Øyvind (Eds.), Developments in Marine Biology. Elsevier Science

Panigada, S., Zanardelli, M., MacKenzie, M., Donovan, C., Melin, F., Hammond, P.S. 2008. Modelling habitat preferences for fin whales and striped dolphins in the Pelagos Sanctuary (Western Mediterranean Sea) with physiographic and remote sensing variables. *Remote Sens. Environ.* 112: 3400-12.

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.

Pearson, R.G., Dawson, T.P., Berry, P.M., Harrison, P.A. 2002. SPECIES: A Spatial Evaluation of Climate Impact on the Envelope of Species. *Ecol. Model.* 154: 289-300.

Pendleton, D.E., Sullivan, P.J., Brown, M.W., Cole, T.V.N., Good , C.P., Mayo, C.A., Monger, B.C., Phillips, S., Record, N.R., Pershing, A. 2012. Weekly predictions of North Atlantic right whale Eubalaena glacialis habitat reveal influence of prey abundance and seasonality of habitat preferences. *Endanger. Species Res.* 18: 147-61.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M., Araújo, M.B. 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton.

Phillips, S.J. and Dudik, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography* 31: 161-75.

Phillips, S.J. and Elith, J. 2013. On estimating probability of presence from use-availability or presence-background data. *Ecology* 94: 1409-19.

Phillips, S.J., Anderson, R.P. and Schapire, R.E. 2006. Maximum entropy modeling of species geographic distributions. *Ecol. Model.* 190: 231-59.

Pourret, O., Naïm, P. and Marcot, B. 2008. *Bayesian networks: A practical guide to applications*. Wiley. 446 pp.

Pouteau, R. Meyer, J-Y., Taputuarai, R. and Stoll, B. 2012. Support vector machines to map rare and endangered native plants in Pacific islands forests. *Ecol. Inform.* 9: 37-46.

Prieto, R., Tobeña, M. and Silva, M.A. 2016. Habitat preferences of baleen whales in a mid-latitude habitat. *Deep Sea Res. II.* doi: 10.1016/j.dsr2.2016.07.015

R Development Core Team 2015. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.

Ramsey, F. L. and Harrison, K. 2004. A closer look at detectability. *Environ. Ecol. Stat.* 11: 73-84.

Raphael, M.G., Wisdom, M.J., Rowland, M.M., Holthausen, R.S., Wales, B.C., Marcot, B.G. and Rich, T.D. 2001. Status and trends of habitats of terrestrial vertebrates in relation to land management in the interior Columbia river basin. *For. Ecol. Manage.* 153: 63-88.

Ready, J., Kaschner, K., South, A.B., Eastwood, P.D., Rees, T., Rius, J., Agbayani, E., Kullander, S. and Froese, R. 2010. Predicting the distributions of marine organisms at the global scale. *Ecol. Model.* 221: 467-78.

Redfern, J.V., Ferguson, M.C., Becker, E.A., Hyrenbach, K.D., Good, C., Barlow, J., Kaschner, K., Baumgartner, M.F., Forney, K.A., Ballance, L.T., Fauchald, P., Halpin, P., Hamazaki, T., Pershing, A.J., Qian, S.S., Read, A., Reilly, S.B., Torres, L., Werner, F. 2006. Techniques for cetacean-habitat modeling. *Mar. Ecol. Prog. Ser.* 310: 271-95.

Renner, M., Parrish, J.K., Piatt, J.F., Kuletz, K.J., Edwards, A.E., Hunt, G.L.J. 2013. Modeled distribution and abundance of a pelagic seabird reveal trends in relation to fisheries. *Mar. Ecol. Prog. Ser.* 484: 259-77.

Ressler, P.H., Dalpadado, P., Macaulay, G.J., Handegard, N. and Skern-Mauritzen, M. 2015. Acoustic surveys of euphausiids and models of baleen whale distribution in the Barents Sea. *Mar. Ecol. Prog. Ser.* 527: 13-29.

Roberts, J.J., Best, B.D., Mannocci, L., Fujioka, E., Halpin, P.N., Palka, D.L., Garrison, L.P., Mullin, K.D., Cole, T.V.N., Khan, C.B., McLellan, W.A., Pabst, D.A. and Lockhart, G.G. 2016. Habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico. *Sci. Rep.* 6: 22615.

Robinson, L.M., Elith, J., Hobday, A.J., Pearson, R.G., Kendall, B.E., Possingham, H.P., Richardson, A.J. 2011. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Glob. Ecol. Biogeogr.* 20: 789-802.

Robertson, M. P., Caithness, N. and Villet, M. H. 2001. A PCA-based modelling technique for predicting environmental suitability for organisms from presence records. *Divers. Distrib.* 7: 15–27.

Royle, J.A., Dorazio, R.M. 2008. *Hierarchical Modeling and Inference in Ecology: The Analysis of Data from Populations, Metapopulations and Communities*. Elsevier Science. 444pp.

Sasaki, H., Murase, H., Kiwada, H., Matsuoka, K., Mitani, Y., Saitoh, S. 2013. Habitat differentiation between sei (*Balaenoptera borealis*) and Bryde's whales (*B.* br*ydei*) *in* the western North Pacific. *Fish. Oceanogr.* 22: 496-508.

Scales, K.L., Miller, P.I., Ingram, S.N., Hazen, E.L., Bograd, S.J., Phillips, R.A. 2015. Identifying predictable foraging habitats for a wide-ranging marine predator using ensemble ecological niche models. *Divers. and Distrib*. 1-13.

Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J. and Williamson, R.C. 2001. Estimating the Support of a High-Dimensional Distribution. Neural Comput. 13: 1443-71.

Scutari, M. and Denis, J.B. 2014. *Bayesian Networks: With Examples in R*. CRC Press. 241pp.

Segurado, P., Araújo, M.B. 2004. An evaluation of methods for modelling species distributions. *J. Biogeogr.* 31: 1555-68.

Shmueli, G. 2010. To Explain or to Predict? *Stat. Sci* 25: 289-310.

Skern-Mauritzen, M., Johannesen, E., Bjørge, A., Øien, N. 2011. Baleen whale distributions and prey associations in the Barents Sea. *Mar. Ecol. Prog. Ser.* 426: 289-301.

Skov, H., Gunnlaugsson, T., Budgell, W.P., Horne, J., Nøttestad, L., Olsen, E., Søiland, H., Víkingsson, G., Waring, G. 2008. Small-scale spatial variability of sperm and sei whales in relation to oceanographic and topographic features along the Mid-Atlantic Ridge. *Deep Sea Res. II* 55: 254-68.

Smith, J.N., Grantham, H.S., Gales, N., Double, M.C., Noad, M.J., Paton, D. 2012. Identification of humpback whale breeding and calving habitat in the Great Barrier Reef. *Mar. Ecol. Prog. Ser.* 447: 259-72.

Stockwell, D.R.B. and Noble, I.R. 1992 Induction of sets of rules from animal distribution data – a robust and informative method of data-analysis. *Math. Comput. Simulat.* 33: 385-390.

Stockwell, D., and Peters, D. 1999. The GARP modelling system: problems and solutions to automated spatial prediction. International Journal of *Geogr. Inf. Sci.* 13: 143-158.

Tantipisanuh, N., Gale, G.A. and Pollino, C. 2014. Bayesian networks for habitat suitability modeling: a potential tool for conservation planning with scarce resources. *Ecol. Appl.* 24: 1705-18.

Tirelli, T., Gamba, M. and Pessani, D. 2012. Support vector machines to model presence/absence of *Alburnus alburnus alborella* (Teleostea, Cyprinidae) in North-Western Italy: comparison with other machine learning techniques. C. R. Biol. 335: 680-6.

Thomas, L., Buckland, S.T., Rexstad, E.A., Laake, J.L., Strindberg, S., Hedley, S.L., Bishop, J.R., Marques, T.A. and Burnham, K.P. 2010. Distance software: design and analysis of distance sampling surveys for estimating population size. *J. Appl. Ecol.* 47: 5-14.

Thomson, J. A., Cooper, A. B., Burkholder, D. A., Heithaus, M. R. and Dill, L. M. 2012. Heterogeneous patterns of availability for detection during visual surveys: spatiotemporal variation in sea turtle dive-surfacing behaviour on a feeding ground. *Methods Ecol. Evol.* 3: 378-387.

Thuiller, W. 2014. Editorial commentary on ‘BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change’. *Glob Change Biol.* 20: 3591–3592.

Thuiller, W., Lafourcade, B., Engler, R., Araújo, M.B., 2009. BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography* 32: 369–373.

Tobeña, M., Prieto, R., Machete, M. and Silva, M.A. 2016. Modeling the potential distribution and richness of cetaceans in the Azores from fisheries observer program data. *Front. Mar. Sci.* 3. doi: 10.3389/fmars.2016.00202

Torres, L.G., Smith, T.D., Sutton, P., MacDiarmid, A., Bannister, J., Miyashita, T. 2013. From exploitation to conservation: habitat models using whaling data predict distribution patterns and threat exposure of an endangered whale. *Diversity and Distributions* 19: 1138-52.

Uusitalo, L. 2007. Advantages and challenges of Bayesian networks in environmental modelling. *Ecol. Model.* 203: 312-8.

Vapnik, V., 1998. Statistical learning theory. Support Vector Machines for Pattern Recognition. John Wiley & Sons, New York.

Víkingsson, G.A., Heide-Jørgensen, M.P. 2015. First indications of autumn migration routes and destination of common minke whales tracked by satellite in the North Atlantic during 2001–2011. *Mar. Mamm. Sci.* 31: 376-85.

Víkingsson, G.A., Pike, D.G., Schleimer, A., Valdimarsson, H., Gunnlaugsson, T., Silva, T., Elvarsson, B.Þ., Mikkelsen, B., Öien, N., Desportes, G., Bogason, V., Hammond, P. 2015. Distribution, abundance and feeding ecology of baleen whales in Icelandic waters: have recent environmental changes had an effect? *Frontiers in Ecology and Evolution* 3.

Verner, J., Morrison, M.L., Ralph, C.J., Pacific Southwest, F., Range Experiment, S., Wildlife, S., San Francisco Bay Area, C. 1986. *Wildlife 2000 : modeling habitat relationships of terrestrial vertebrates : based on an international symposium held at Stanford Sierra Camp, Fallen Leaf Lake, California, 7-11 October 1984*. University of Wisconsin Press, Madison, Wis.

Wheeler, B., Gilbert, M., Rowe, S. 2012. Definition of critical summer and fall habitat for bowhead whales in the eastern Canadian Arctic. *Endanger. Species Res.* 17: 1-16.

Williams, R., Hedley, S.L., Branch, T.A., Bravington, M.V., Zerbini, A.N., Findlay, K.P. 2011. Chilean blue whales as a case study to illustrate methods to estimate abundance and evaluate conservation status of rare species. *Conservation biology : the journal of the Society for Conservation Biology* 25: 526-35.

Williams, R., Hedley, S.L., Hammond, P.S. 2006. Modeling distribution and abundance of Antarctic baleen whales using ships of opportunity. *Ecol. Soc.* 11.

Williams, R., Kelly, N., Boebel, O., Friedlaender, A.S., Herr, H., Kock, K.H., Lehnert, L.S., Maksym, T., Roberts, J., Scheidat, M., Siebert, U., Brierley, A.S. 2014. Counting whales in a challenging, changing environment. *Scientific Reports* 4.

Wood, S., N. 2006. *Generalized additive models: an introduction with R.* Chapman & Hall/CRC, Boca Raton.

Yackulic, C. B., Chandler, R., Zipkin, E. F.. Royle, J. A., Nichols, J. D., Grant, E. H. C. and Veran, S. 2013. Presence-only modelling using MAXENT: when can we trust the inferences? *Methods Ecol. Evol.* 4: 236-43.

Zerbini, A.N., Friday, N.A., Palacios, D.M., Waite, J.M., Ressler, P.H., Rone, B.K., Moore, S.E. and Clapham, P.J. 2016. Baleen whale abundance and distribution in relation to environmental variables and prey density in the Eastern Bering Sea. *Deep Sea Res. II* 134: 312-30.

Table 1. Classification of species distribution models (SDMs) based of modelling methods and response variables.

Type	Response variables		
	Presence/absence	Presence/background	Presence only
Regression	GLM GAM	-	-
Profile	-	-	HSI BIOCLIM DOMAIN
Machine learning	CART BRT RF GA SVM	MaxEnt	-
Other	HBM	ENFA	-

Table 2. Some examples of statistical models which can deal with spatial autocorrelation (SAC) and other correlation structures.

Name of model	Abbreviation
Autocovariate models	-
Spatial eigenvector mapping	SEVM
Generalized least squares	GLS
Conditional autoregressive model	CAR
Simultaneous autoregressive model	SAR
Generalized linear mixed model	GLMM
Generalized estimating equations	GEE

Table 3. Target species, regions and areas of species distribution models (SDMs) applied to baleen whales. Papers published in scientific journals from December 1997 to December 2016 are considered. It should be noted that some papers dealt multi-species and/or multi-regions.

Common name	Scientific name		Region	Area	Reference
Bowhead whale	*Balaena*	*mysticetus*	Arctic	Eastern Canada	Wheeler et al. (2012)
North Atlantic right whale	*Eubalaena*	*glacialis*	North Atlantic	-	Kaschner et al. (2006)
				-	Monsarrat, et al. (2015)
				-	Monsarrat, et al. (2016)
				Cape Cod Bay	Pendleton et al. (2012)
				Coast of Florida and Georgia	Keller et al. (2012)
				Nova Scotian Shelf	Moses and Finn (1997)
				US east coast	Best et al. (2012)
North Pacific right whale	*Eubalaena*	*japonica*	North Pacific	-	Gregr (2011)
				-	Monsarrat, et al. (2015)
				-	Monsarrat, et al. (2016)
Southern right whale	*Eubalaena*	*australis*	Southern Ocean	Australasian region	Torres et al. (2013)
Gray whale	*Eschrichtius*	*robustus*	North Pacific	-	Kaschner et al. (2006)
Common minke whale	*Balaenoptera*	*acutorostrata*	Arctic	Barents Sea	Ressler (2015)
					Skern-Mauritzen et al. (2011)
			Bering Sea	Eastern Bering Sea	Zerbini (2016)
			North Atlantic	Azores islands	Tobeña et al. (2016)
				US east coast	Hamazaki (2002)
					Robert (2016)
			North Pacific	British Columbia	Best et al. (2015)
				Western North Pacific	Okamura et al. (2001)
Antarctic minke whale	*Balaenoptera*	*bonaerensis*	Antarctic	0°-40°E	Hedley et al. (1999)
				140°E-35°W	Ainley et al. (2012)
				60°E-60°W	Bombosch et al. (2014)
				65°W-55°W	Williams et al. (2006)
				Circumpolar	Beekmans et al. (2010)
					Kaschner et al. (2006)
				Marguerite Bay	Friedlaender et al. (2011)
				Ross Sea	Ballard et al. (2012)
					Murase et al. (2013)
				Weddell Sea	Williams et al. (2014)
Sei whale	*Balaenoptera*	*borealis*	North Atlantic	Azores islands	Prieto et al. (2016)
				Azores islands	Tobeña et al. (2016)
				Mid-Atlantic Ridge	Skov et al. (2008)
				US east coast	Robert (2016)
			North Pacific	British Columbia	Gregr and Trites (2001)
				Western North Pacific	Murase et al. (2014)
					Sasaki et al. (2013)
Bryde’s whale	*Balaenoptera*	*edeni*	Tropical Pacific	Eastern tropical Pacific	Forney et al. (2012)
			Gulf of Mexico	US Gulf of Mexico	Robert (2016)
			North Atlantic	US east coast	Robert (2016)
			North Pacific	California Current System	Forney et al. (2012)
				Central North Pacific	Forney et al. (2015)
				Western North Pacific	Sasaki et al. (2013)

Table 3. (continue)

Common name	Scientific name		Region	Area	Reference
Bowhead whale	*Balaena*	*mysticetus*	Arctic	Eastern Canada	Wheeler et al. (2012)
Blue whale	*Balaenoptera*	*musculus*	North Atlantic	Azores islands	Prieto et al. (2016)
					Tobeña et al. (2016)
				US east coast	Robert (2016)
			North Pacific	British Columbia	Gregr and Trites (2001)
				California Current System	Becker et al. (2015)
					Forney et al. (2012)
				Eastern tropical Pacific	Forney et al. (2012)
			Pacific Ocean	East Pacific	Pardo et al. (2015)
			South Pacific	Coast of Chile	Williams et al. (2011)
Fin whale	*Balaenoptera*	*physalus*	Antarctic	65°W-55°W	Williams et al. (2006)
				Circumpolar	Ready et al. (2010)
				West Antarctic Peninsula	Herr et al. (2016)
			Arctic	Barents Sea	Ressler (2015)
					Skern-Mauritzen et al. (2011)
				West coast of Greenland	Laidre et al. (2010)
			Atlantic	Island	Vikingsson et al. (2015)
			Bering Sea	Eastern Bering Sea	Zerbini (2016)
			Gulf of Mexico	US Gulf of Mexico	Robert (2016)
			Mediterranean Sea	Northwestern Mediterranean Sea	Laran and Gannier (2008)
				Pelagos Sanctuary	Panigada et al. (2008)
				Western Mediterranean Sea	Cotté et al. (2009)
			North Atlantic	Azores islands	Prieto et al. (2016)
					Tobeña et al. (2016)
				US east coast	Hamazaki (2002)
					Robert (2016)
			North Pacific	British Columbia	Best et al. (2015)
					Gregr and Trites (2001)
				California Current System	Becker et al. (2012)
					Becker et al. (2015)
					Forney et al. (2012)
Humpback whale	*Megaptera*	*novaeangliae*	Antarctic	60°E-60°W	Bombosch et al. (2014)
				65°W-55°W	Williams et al. (2006)
				Marguerite Bay	Friedlaender et al. (2011)
				West Antarctic Peninsula	Herr et al. (2016)
			Arctic	Barents Sea	Ressler (2015)
				Barents Sea	Skern-Mauritzen et al. (2011)
				West coast of Greenland	Laidre et al. (2010)
			Bering Sea	Eastern Bering Sea	Zerbini (2016)
			Indian Ocean	Arabian Sea off Oman	Corkeron et al. (2011)
			North Atlantic	US east coast	Hamazaki (2002)
					Best et al. (2012)
					Robert (2016)
			North Pacific	British Columbia	Best et al. (2015)
					Dalla Rosa et al. (2012)
					Gregr and Trites (2001)
				California Current System	Becker et al. (2015)
					Forney et al. (2012)
			South Pacific	Great Barrier Reef	Smith et al. (2012)

Table 4. Type of cetacean data and response variables used in species distribution models (SDMs) applied to baleen whales. Papers published in scientific journals from December 1997 to December 2016 are considered.

Reference	Cetacean data	Response variable
Ainley et al. (2012)	Opportunistic sighting	Presence only
Ballard et al. (2012)	Dedicated and opportunistic sighting	Presence only
Becker et al. (2012)	Dedicated sighting	Count
Becker et al. (2015)	Dedicated sighting	Count
Beekmans et al. (2010)	Dedicated sighting	Count
Best et al. (2012)	Dedicated sighting	Presence/absence
Best et al. (2015)	Dedicated sighting	Count
Bombosch et al. (2014)	Opportunistic sighting	Presence only
Corkeron et al. (2011)	Opportunistic sighting	Count
Cotté et al. (2009)	Opportunistic sighting	Count
Dalla Rosa et al. (2012)	Opportunistic sighting	Count
Forney et al. (2012)	Dedicated sighting	Count
Forney et al. (2015)	Dedicated sighting	Count
Friedlaender et al. (2011)	Opportunistic sighting	Presence only
Gregr (2011)	Catch	Presence only
Gregr and Trites (2001)	Catch	Count
Hamazaki (2002)	Dedicated sighting	Presence/absence
Hedley et al. (1999)	Dedicated sighting	Count
Herr et al. (2016)	Dedicated sighting	Count
Kaschner et al. (2006)	Published data	-
Keller et al. (2012)	Dedicated sighting	Count
Laidre et al. (2010)	Dedicated sighting	Presence only
Laran and Gannier (2008)	Opportunistic sighting	Presence/absence
Monsarrat, et al. (2015)	Catch	Presence/absence
Monsarrat, et al. (2016)	Catch	Count
Moses and Finn (1997)	Dedicated sighting	Presence/absence
Murase et al. (2013)	Dedicated sighting	Count
Murase et al. (2014)	Dedicated sighting	Count
Okamura et al. (2001)	Dedicated sighting	Count
Panigada et al. (2008)	Opportunistic sighting	Presence/absence
Pardo et al. (2015)	Dedicated sighting	Count
Pendleton et al. (2012)	Dedicated sighting	Presence only
Prieto et al. (2016)	Opportunistic sighting	Presence only
Ready et al. (2010)	Dedicated sighting	Presence/absence
Ressler (2015)	Opportunistic sighting	Count
Robert (2016)	Dedicated sighting	Count
Sasaki et al. (2013)	Dedicated sighting	Presence/absence
Skern-Mauritzen et al. (2011)	Opportunistic sighting	Count
Skov et al. (2008)	Opportunistic sighting	Presence only
Smith et al. (2012)	Opportunistic sighting	Presence only
Tobeña et al. (2016)	Opportunistic sighting	Presence only
Torres et al. (2013)	Catch	Presence/absence
Vikingsson et al. (2015)	Dedicated sighting	Count
Wheeler et al. (2012)	Various (e.g. sighting and catch)	Presence only
Williams et al. (2006)	Opportunistic sighting	Count
Williams et al. (2011)	Dedicated sighting	Count
Williams et al. (2014)	Dedicated sighting	Count
Zerbini (2016)	Opportunistic sighting	Count

Table 5. Explanatory variables used in species distribution models (SDMs) applied to baleen whales. Papers published in scientific journals from December 1997 to December 2016 are considered. SST: sea surface temperature; SSH: sea surface height (including its anomaly); Chl: sea surface chlorophyll *a* concentrations; Current: sea surface current; SSS: sea surface salinity.

Reference	Latitude	Longitude	Year	Month	Distance to shore	Depth	Slope	Aspect	Bottom complexity	Distance from bottom terrain (e.g. shelf)	Seamount density/ depth
Ainley et al. (2012)	-	-	-	-	-	X	X	-	-	-	
Ballard et al. (2012)	-	-	-	-	-	X	X	-	-	-	
Becker et al. (2012)	-	-	-	-	-	X	X	-	-	-	
Becker et al. (2015)	X	X	X	-	-	X	X	X	-	-	
Beekmans et al. (2010)	X	X	-	-	-	X	-	-	-	X	
Best et al. (2012)	-	-	-	-	X	X	-	-	-	-	
Best et al. (2015)	X	X	-	-	X	X	X	-	-	-	
Bombosch et al. (2014)	-	-	-	-	-	X	X	-	-	-	
Corkeron et al. (2011)	-	-	-	-	X	X	X		-	-	
Cotté et al. (2009)	-	-	-	X	X	X	-	-	-	-	
Dalla Rosa et al. (2012)	X	X	X	X	X	X	X		-	X	
Forney et al. (2012)	-	-	-	-	X	X	X	-	-	X	
Forney et al. (2015)	X	X	-	-	X	-	-	-	-	-	
Friedlaender et al. (2011)	-	-	-	-	X	X	X	-	-	-	
Gregr (2011)	-	-	-	-	-	X	-	-	-	-	
Gregr and Trites (2001)	X	X	X	X	-	X	X	-	-	-	
Hamazaki (2002)	-	-	-	-		X	X	-	-	-	
Hedley et al. (1999)	X	X	-	-	-	-	-	-	-	-	
Herr et al. (2016)	X	X	-	-	-	X	-	-	-	-	
Kaschner et al. (2006)	-	-	-	-	-	X	-	-	-	-	
Keller et al. (2012)	-	-	-	-	-	X	-	-	-	-	
Laidre et al. (2010)	-	-	-	-	-	-	-	-	-	-	
Laran and Gannier (2008)	-	-	-	-	-	X	-	-	-	X	
Monsarrat, et al. (2015)	-	-	-	-	X	X	X	-	-	-	
Monsarrat, et al. (2016)	-	-	-	-	X	X	X	-	-	-	
Moses and Finn (1997)	-	-	-	-	-	X	-	-	-	-	
Murase et al. (2013)	X	X	-	-	-	X	-	-	-	X	
Murase et al. (2014)	-	X	-	-	-	-	-	-	-	-	
Okamura et al. (2001)	X	X	X	X	-	-	-	-	-	-	
Panigada et al. (2008)	X	X	X	-	X	X	X	-	-	-	
Pardo et al. (2015)	-	-	-	-	-	-	-	-	-	-	
Pendleton et al. (2012)	-	-	-	-	-	X	-	-	-	-	
Prieto et al. (2016)	-	-	-	-	X	X	X	-	-	X	X
Ready et al. (2010)	-	-	-	-	-	X	-	-	-	-	
Ressler (2015)	X	X	X	-	-	X	-	-	-	-	
Robert (2016)	-	-	-	-	X	X	X	-	-	X	-
Sasaki et al. (2013)	-	-	-	-	-	X	-	-	-	-	
Skern-Mauritzen et al. (2011)	X	X	X	-	-	-	-	-	-	-	
Skov et al. (2008)	-	-	-	-	-	X	X	-	X	-	
Smith et al. (2012)	-	-	-	-	X	X	X		-	X	
Tobeña et al. (2016)	-	-	-	-	X	X	X	-	-	X	X
Torres et al. (2013)	-	-	-	-	-	X	X	-	-	X	
Vikingsson et al. (2015)	-	-	-	-	X	X	X	X	-	X	
Wheeler et al. (2012)	-	-	-	-	X	X	X	-	-	-	
Williams et al. (2006)	X	X	-	-	X	X	-	-	-	-	
Williams et al. (2011)	X	X	-	-	-	-	-	-	-	-	
Williams et al. (2014)	X	X	-	-	-	-	-	-	-	-	
Zerbini (2016)	-	-	-	-	X	X	X	-	-	X	-

Table 5 (continue)

Reference	SST	SSH	TKE EKE	Distance to eddy	Chl/ PP	Current	SSS	Subsurface temperature	Subsurface salinity	MLD
Ainley et al. (2012)	-	-			X	-	-	-	-	-
Ballard et al. (2012)	-	-			X	-	-	-	-	-
Becker et al. (2012)	X	-			-	-	-	-	-	-
Becker et al. (2015)	X	X			X	-	X	-	-	X
Beekmans et al. (2010)	X	-			X	X	-	-	-	-
Best et al. (2012)	X	X			X	-	-	-	-	-
Best et al. (2015)	-	-			-	-	-	-	-	-
Bombosch et al. (2014)	X	X			X	-	-	-	-	-
Corkeron et al. (2011)	-	-			-	-	-	-	-	-
Cotté et al. (2009)	X	X			X	-	X	-	-	-
Dalla Rosa et al. (2012)	X	X			X	X	X	X	X	-
Forney et al. (2012)	X	-			X	-	X	-	-	X
Forney et al. (2015)	X	X			X	-	-	-	-	-
Friedlaender et al. (2011)	-	-			-	-	-	X	-	-
Gregr (2011)	X	-			-	X	-	X	-	
Gregr and Trites (2001)	X	-			-	-	X	-	-	-
Hamazaki (2002)	X	-			-	-	-	-	-	-
Hedley et al. (1999)	-	-			-	-	-	-	-	-
Herr et al. (2016)	-	-			-	-	-	X	X	-
Kaschner et al. (2006)	X	-			-	-	-	-	-	-
Keller et al. (2012)	X	-			-	-	-	-	-	-
Laidre et al. (2010)	-	-			-	-	-	-	-	-
Laran and Gannier (2008)	X	-			X	-	-	-	-	-
Monsarrat, et al. (2015)	X	X			X	-	-	-	-	X
Monsarrat, et al. (2016)	X	X			X	-	-	-	-	X
Moses and Finn (1997)	X	-			-	-	-	-	-	-
Murase et al. (2013)	-	-			-	-	-	X	X	-
Murase et al. (2014)	-	-			-	-	-	-	-	-
Okamura et al. (2001)	X	-			-	-	-	-	-	-
Panigada et al. (2008)	X	-			X	-	-	-	-	-
Pardo et al. (2015)	-	X			-	-	-	-	-	-
Pendleton et al. (2012)	X	-			X	-	-	-	-	-
Prieto et al. (2016)	X	-			X	-	-	-	-	-
Ready et al. (2010)	X	-			X	-	-	-	-	-
Ressler (2015)	-	-			-	-	-	-	-	-
Robert (2016)	X	-	X	X	X	-	-	-	.	-
Sasaki et al. (2013)	X	X			X	-	-	-	-	-
Skern-Mauritzen et al. (2011)	-	-			-	-	-	X	-	-
Skov et al. (2008)	-	X			-	X	-	X	X	-
Smith et al. (2012)	X	-			-	-	-	-	-	-
Tobeña et al. (2016)	X	-			X	-	-	-	-	-
Torres et al. (2013)	X	-			X	-	-	X	-	X
Vikingsson et al. (2015)	X	X			-	-	-	-	-	-
Wheeler et al. (2012)	X	-			X	-	-	-	-	-
Williams et al. (2006)	-	-			-	-	-	-	-	-
Williams et al. (2011)	-	-			-	-	-	-	-	-
Williams et al. (2014)	-	-			-	-	-	-	-	-
Zerbini (2016)	X	-	-	-	X	-	-	-	-	-

Table 5 (continue)

Reference	Thermal front	Oceanic front	Water mass	Distance from sea ice	Sea ice cover	Sea ice concentration	Prey	Wind speed/ Beaufort	Visibility	Length of day
Ainley et al. (2012)	-	X	-	-	X	-	-	-		-
Ballard et al. (2012)	-	X	X	-	X	-	-	-		-
Becker et al. (2012)	-	-	-	-	-	-	-	X		-
Becker et al. (2015)	-	-	-	-	-	-	-	-		-
Beekmans et al. (2010)	-	X	-	X	-	-	-	-		-
Best et al. (2012)	-	-	-	-	-	-	-	-		-
Best et al. (2015)	-	-	-	-	-	-	-	-		-
Bombosch et al. (2014)	-	-	-	X	-	X	-	-		X
Corkeron et al. (2011)	-	-	-	-	-	-	-	-		-
Cotté et al. (2009)	-	-	-	-	-	-	-	-		-
Dalla Rosa et al. (2012)	X	-	-	-	-	-	-	-		-
Forney et al. (2012)	-	-	-	-	-	-	-	X		-
Forney et al. (2015)	-	-	-	-	-	-	-	-		-
Friedlaender et al. (2011)	-	-	-	X	-	-	X	-		-
Gregr (2011)	-	-	-		-	-	-			-
Gregr and Trites (2001)	-	-	-	-	-	-	-	-		-
Hamazaki (2002)	X	-	-	-	-	-	-	-		-
Hedley et al. (1999)	-	-	-	X	-	-	-	-		-
Herr et al. (2016)	-	-	-	-	-	X	-	-		-
Kaschner et al. (2006)	-	-	-	X	-	-	-	-		-
BesBes et al. (2012)	-	-	-	-	-	-	-	X		-
Laidre et al. (2010)	-	-	-	-	-	-	-	-		-
Laran and Gannier (2008)	-	-	-	-	-	-	-	-		-
Monsarrat, et al. (2015)	-	-	-	-	-	-	-	-		-
Monsarrat, et al. (2016)	-	-	-	-	-	-	-	-		-
Moses and Finn (1997)	-	-	-	-	-	-	-	-		-
Murase et al. (2013)	-	-	-	-	-	-	X	-		-
Murase et al. (2014)	-	X	-	-	-	-	-	-		-
Okamura et al. (2001)	-	-	-	-	-	-	-	-		-
Panigada et al. (2008)	-	-	-	-	-	-	-	-		-
Pardo et al. (2015)	-	-	-	-	-	-	-	-		-
Pendleton et al. (2012)	-	-	-	-	-	-	X	-		-
Prieto et al. (2016)	-	-	-	-	-	-	-	-		-
Ready et al. (2010)	-	-	-	-	-	-	-	-		-
Ressler (2015)	-	-	-	-	-	-	X	X	X	-
Robert (2016)	-	-	-	-	-	-	X	X	-	-
Sasaki et al. (2013)	-	-	-	-	-	-	-	-		-
Skern-Mauritzen et al. (2011)	-	-	-	-	-	-	-	-		-
Skov et al. (2008)	-	-	-	-	-	-	-	-		-
Smith et al. (2012)	-	-	-	-	-	-	-	-		-
Tobeña et al. (2016)	-	-	-	-	-	-	-	-		-
Torres et al. (2013)	-	-	-	-	-	-	-	-		-
Vikingsson et al. (2015)	-	-	-	-	-	-	-	-		-
Wheeler et al. (2012)	-	-	-	X	-	-	-	-		-
Williams et al. (2006)	-	-	-	-	-	-	-	-		-
Williams et al. (2011)	-	-	-	-	-	-	-	-		-
Williams et al. (2014)	-	-	-	X	-	X	-	-		-
Zerbini (2016)	-	-	-	-	-	-	X	-	-	-

Table 6. Summary of species distribution models (SDMs) applied to baleen whales. Papers published in scientific journals from December 1997 to December 2016 are considered.

Reference	Resolution for spatial estimation	Segment size	Model	Error distribution	Model evaluation	Spatial autocorrelation	Collinearity
Ainley et al. (2012)	5 km	-	MaxEnt	-	AUC	-	-
Ballard et al. (2012)	5 km	-	MaxEnt	-	AUC	-	-
Becker et al. (2012)	25 km	5 km	GAM	Quasi-Poisson	Visual comparison Comparison with DISTANCE estimate Spearman rank correlation test	-	-
Becker et al. (2015)	10km	5×5.5km	GAM	Tweedie	Deviance explained RMSE Ratio observation vs estimate Comparison with DISTANCE estimate	-	Method unknown
Beekmans et al. (2010)	0.2°	10 n,miles	GAM	Tweedie Quasi-Poisson Poison	-	-	-
Best et al. (2012)	10 km	-	GAM	Quasi-bionomial	AUC	-	-
Best et al. (2015)	5 km	1 n.mile	GAM	Quasi-Poisson	Deviance explained	-	-
Bombosch et al. (2014)	0.25°	-	MaxEnt	-	AUC	-	-
Corkeron et al. (2011)	0.1°	-	GLM	Quasi-Poisson	-	SEVM	-
Cotté et al. (2009)	NA	-	GAM	Gamma	-	-	-
Dalla Rosa et al. (2012)	4.63 km	4 km	GAM	Quasi-Poisson	-	Variogram	-
Forney et al. (2012)	25 km	10 km	GAM	Quasi-Poisson	-	-	-
Forney et al. (2012)	25 km	5 km	GAM	Quasi-Poisson	-	-	-
Forney et al. (2015)	25 km	10 km	GAM	Quasi-Poisson	Deviance explained Average squared prediction error Comparison with DISTANCE estimate	Moran’s I	Method unknown
Friedlaender et al. (2011)	1 km	-	MaxEnt	-	AUC	-	-
Gregr (2011)	50 km	-	MaxEnt		AUC		-
Gregr and Trites (2001)	10 km	-	GLM	Poisson	Cross validation Classification tables	-	-
Hamazaki (2002)	10'	-	Logistic regression	Binomial	-	-	-
Hedley et al. (1999)	-	5 n.miles	GAM	Poisson	Comparison with DISTANCE estimate	-	-
Herr et al. (2016)	6.25 km	31.75 km (avg.)	GAM	Negative binomial	Deviance explained	-	-
Kaschner et al. (2006)	0.5°	-	RES	-	Comparison with actual data	-	-
Keller et al. (2012)	4 km	-	GAM	Poisson	Bootstrap	GLMM	-
Laidre et al. (2010)	2 km	-	Kernel method	-	-	-	-
Laran and Gannier (2008)	10 n.miles	-	Logistic regression	Binomial	AUC	-	-

Table 6 (continue)

Reference	Resolution for spatial estimation	Segment size	Model	Error distribution	Model evaluation	Spatial autocorrelation	Collinearity
Monsarrat, et al. (2016)	1°	-	GAM	Negative binomial Poisson	Deviance explained K-fold Cross validation	-	-
Moses and Finn (1997)	10'	-	Logistic regression	Binomial	Visual comparison	-	-
Murase et al. (2013)	10 km	-	GAM	Poisson	Comparison with DISTANCE estimate	-	-
Murase et al. (2014)	30 km	-	GAM	Tweedie	Comparison with DISTANCE estimate	-	VIF
Okamura et al. (2001)	1°	-	GAM	Poisson	-	-	-
Panigada et al. (2008)	2'	-	GAM	Binomial	-	-	-
Pardo et al. (2015)	1/3°	-	Hierarchical Bayesian	-	-	-	-
Pendleton et al. (2012)	1 km	-	MaxEnt	-	AUC	-	-
Prieto et al. (2016)	NA	-	MaxEnt	-	K-fold cross validation AUC	-	Kendeall's correlation coefficient
Ready et al. (2010)	0.5°	-	RES	-	AUC Comparison with actual data	-	-
			GARP	-			
			GLM	Binomial			
			GAM	Binomial			
			MaxEnt	-			
Ressler (2015)	10, 50 km	-	GAMM	Poisson	Deviance explained	-	-
Robert (2016)	10 km	10 km	GAM	Tweedie	-	-	-
Sasaki et al. (2013)	4 km	-	GLM	Binomial	-	Moran’s I	-
Skern-Mauritzen et al. (2011)	50 km	-	GAMM	Quasi-Poisson	-	-	VIF
Skov et al. (2008)	1 km	-	ENFA	-	-	-	-
Smith et al. (2012)	4.8 km	-	MaxEnt	-	AUC	-	-
Tobeña et al. (2016)	NA	-	MaxEnt	-	TSS AUC	-	
Torres et al. (2013)	25 km	-	BRT	-	AUC	-	-
Vikingsson et al. (2015)	0.5°	1.75 n.mile	GAM	Negative binomial	-	Variogram	Pearson Correlation coefficient
Wheeler et al. (2012)	10 km	-	ENFA	-	Jack-knife	-	-
Williams et al. (2006)	2 n.miles	5 km	GAM	Quasi-Poisson	-	-	-
Williams et al. (2011)	NA	20 n.miles	GAM	Tweedie	-	-	-
Williams et al. (2014)	6.25 km	30 n.miles	GAM	Tweedie	-	-	-
Zerbini (2016)	9 km	9 km	GAM	Poisson	Deviance explained	-	Pearson Correlation coefficient

Table 7. Summary of generalized additive models (GAMs) applied to baleen whales as species distribution models (SDMs). Papers published in scientific journals from December 1997 to December 2016 are considered.

Reference	Software	Version	Package	Version	Error distribution	Smoother	Dimension parameter (K)	Gamma	Power parameter (P)	Model selection
Becker et al. (2012)	S-PLUS	6.1 Release 1	-	-	Quasi-Poisson	NA	NA	NA	-	AIC
Becker et al. (2015)	R	3.1.1	mgcv	1.8-3	Tweedie	Thin-plate Tensor	NA	NA	NA	REML
Beekmans et al. (2010)	R	NA	mgcv	1.5-5	Tweedie Quasi-Poisson Poison	Isotropic Tensor	NA	1.4	NA	GCV
Best et al. (2012)	R	NA	mgcv	NA	Quasi-binomial	Thin-plate	NA	1.4	-	GCV
Best et al. (2015)	R	NA	mgcv	NA	Quasi-Poisson	Thin-plate	5	NA	-	GCV
Cotté et al. (2009)	R	NA	mgcv	NA	Gamma	NA	NA	NA	-	GCV
Dalla Rosa et al. (2012)	R	NA	mgcv	1.4-1	Quasi-Poisson		8	1.4	-	GCV
Forney et al. (2012)	S-PLUS R R	NA 2.6.2 2.6.2	- gam mgcv	- NA 1.3-29	NA	Cubic spline Cubic spline Cubic and thin-plate spline	3 3 NA	1.0 & 1.4	-	AIC AIC GCV
Forney et al. (2012)	S-PLUS R R	NA 2.6.2 2.6.2	- gam mgcv	- NA 1.3-29	NA	Cubic spline Cubic spline Cubic and thin-plate spline	3 3 NA	1.0 & 1.4	-	AIC AIC GCV
Forney et al. (2015)	Spotfire S+	8.1	-	-	Quasi-Poisson	Thin-plate	3	NA	-	AIC
Hedley et al. (1999)	NA	NA	NA	NA	Overdispersed-Poisson	Cubic spline	2, 4, 8	NA	-	AIC
Herr et al. (2016)	R	NA	dsm	NA	Negative binomial	NA	-	-	-	UBRE
Keller et al. (2012)	NA	NA	NA	NA	Poisson	NA	2	NA	-	AIC
Monsarrat, et al. (2016)	NA	NA	NA	NA	Negative binomial Poisson	NA	NA	NA	-	AIC
Murase et al. (2013)	R	2.12.1	mgcv	1.7-2	Poisson	NA	NA	NA	-	GCV
Murase et al. (2014)	R	3.0.2	mgcv	1.7-26	Tweedie	NA	NA	NA	1.1	GCV
Okamura et al. (2001)	S-PLUS	NA	-	-	Poisson	NA	NA	NA	-	BIC
Ressler (2015)	R	2.15.2	gamm4	NA	Poisson	NA	NA	NA	-	AIC
Robert (2016)	R	1.8-4	dsm	2.25	Tweedie	Thin-plate	NA	NA	Estimated	REML
Skern-Mauritzen et al. (2011)	R	2.7.0	mgcv	NA	Quasi-Poisson	NA	NA	NA	-	
Vikingsson et al. (2015)	R	NA	mgcv	1.8-1	Negative binomial	Cubic regression	NA	NA	-	GCV
Williams et al. (2006)	R	NA	mgcv	NA	Overdispersed-Poisson	NA	NA	NA	-	GCV
Williams et al. (2011)	R	NA	mgcv	NA	Tweedie	NA	NA	NA	1.1	REML
Williams et al. (2014)	R	NA	mgcv	NA	Tweedie	Tensor product	NA	NA	1.6	REML
Zerbini (2016)	R	NA	mgcv	1.7-26	Poission	NA	Changed	1.4	-	GCV