

Predicting the distributions of marine organisms at the global scale

Jonathan Ready^{a,*}, Kristin Kaschner^b, Andy B. South^c, Paul D. Eastwood^c, Tony Rees^d, Josephine Rius^e, Eli Agbayani^e, Sven Kullander^f, Rainer Froese^g

^a Instituto de Estudos Costeiros, Universidade Federal do Pará – Campus de Bragança, Aldeia, Bragança 68600-000, Pará, Brazil

^b Evolutionary Biology & Ecology Lab, Institute of Biology I (Zoology), Albert-Ludwigs-University Freiburg, Germany

^c Centre for Environment, Fisheries, and Aquaculture Science (Cefas), Lowestoft Laboratory, Lowestoft, Suffolk, NR33 0HT, UK

^d CSIRO Marine and Atmospheric Research, GPO Box 1538 Hobart TAS 7001, Australia

^e WorldFish Center – Philippines Office, MCPO Box 2631, 0718 Makati City, Philippines

^f Department of Vertebrate Zoology, Swedish Museum of Natural History, Box 50007, 104 05 Stockholm, Sweden

^g Leibniz-Institute of Marine Sciences, Düsterbrookweg 20, D-24105 Kiel, Germany

ARTICLE INFO

Article history:

Received 20 March 2008

Received in revised form

15 September 2009

Accepted 16 October 2009

Available online 24 November 2009

Keywords:

Species distribution modelling

Range maps

Global marine biodiversity

Trawl surveys

Expert review

Model comparison

ABSTRACT

We present and evaluate AquaMaps, a presence-only species distribution modelling system that allows the incorporation of expert knowledge about habitat usage and was designed for maximum output of standardized species range maps at the global scale. In the marine environment there is a significant challenge to the production of range maps due to large biases in the amount and location of occurrence data for most species. AquaMaps is compared with traditional presence-only species distribution modelling methods to determine the quality of outputs under equivalently automated conditions. The effect of the inclusion of expert knowledge to AquaMaps is also investigated. Model outputs were tested internally, through data partitioning, and externally against independent survey data to determine the ability of models to predict presence versus absence. Models were also tested externally by assessing correlation with independent survey estimates of relative species abundance. AquaMaps outputs compare well to the existing methods tested, and inclusion of expert knowledge results in a general improvement in model outputs. The transparency, speed and adaptability of the AquaMaps system, as well as the existing online framework which allows expert review to compensate for sampling biases and thus improve model predictions are proposed as additional benefits for public and research use alike.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Concerns over changing patterns of marine biodiversity resulting from climate change and human impacts have generated considerable interest in the use of models designed to generate spatial predictions (i.e. maps) of species' distributions from point occurrence data (Guisan and Thuiller, 2005). Ideally, prediction models would be generated from comprehensive species occurrence and absence data from targeted surveys. Unfortunately this level of data is only available for a relatively limited number of species and geographic locations, creating problems for assessments of changes in patterns of marine species distributions and biodiversity at regional and global scales.

As an alternative, modellers are making use of increasing volumes of presence-only data (Pearce and Boyce, 2006). These are published online through global databases such as FishBase (Froese

and Pauly, 2007) and the Ocean Biogeographic Information System (OBIS, 2007), both of which feed data directly into the Global Biodiversity Information Facility (GBIF, 2007). These data frameworks compile species occurrence data from museum records and other sources. They therefore represent a highly patchy and biased view of patterns of species' distributions as a result of regional and local variations in sampling effort. The bias inherent to the data creates problems when data-driven modelling techniques are used to generate predictions of species' distributions. This is because an absence of occurrence records may not necessarily indicate a true absence in the distribution of the species, but rather a lack of adequate sampling. This is especially true for marine organisms, as inshore areas are more often sampled at a higher rate compared to offshore areas, causing a bias in the species–habitat relationship described by the data (Kaschner et al., 2006; MacLeod et al., 2008). In this scenario, an offshore species might well be predicted to have an inshore distribution if sampling had only occurred over a limited proportion of its overall depth range. Similarly, misidentification of species is a common weakness of all existing large online occurrence record deposi-

* Corresponding author. Tel.: +55 0 91 3425 1745; fax: +55 0 91 3425 1209.
E-mail address: jonathan.ready@gmail.com (J. Ready).

taries (Meier and Dikow, 2004), which in turn can lead to false predicted presences and unrealistic species distribution if these data sets are used as input for standard species distribution modelling.

Until better data sets are available, these biases in sampling effort can be best countered if model algorithms are able to incorporate expert information on species–habitat preferences. These represent a rich but currently underutilized resource. Here, we define expert information as habitat use information that is not directly available as raw data, i.e. published information about habitat use/preference that is based on quantitative investigations of species occurrence in relation to environmental knowledge. Examples include: evidence of a pelagic lifestyle, known depth ranges, latitudinal and longitudinal limits to ranges or physiological tolerances of species. Additionally, experts working on the taxa could include personal knowledge either about occurrence records not yet accessible through online data depositaries, or maximum range extents not described in the literature. This information could also be included should such experts review a map. However, as research into species distribution modelling has progressed, so has the complexity of model algorithms to the extent that users have little or no opportunity to influence the model outcome through the use of expert information. While the goal of species distribution modelling is to increase prediction accuracy (which might be expected to increase with model complexity), the use of increasingly sophisticated methods may also be a barrier to non-expert modellers such as biodiversity managers, decision makers, and planners. All of these people have a vested interest in the reliability of model outputs and therefore need to understand how the models were constructed. Simple and transparent numerical approaches combined with expert guidance on the form of the species–habitat relationships may therefore help circumvent some of the inherent problems in predicting regional and global distributions from patchy, heavily biased occurrence data from global biodiversity databases. If these same algorithms are transparent and produce reliable and verifiable results, the likelihood that predictions will have practical use and feed into decision making and planning will be further increased.

We describe such an approach, called AquaMaps (available for use via the webpage <http://www.aquamaps.org>, and based on the global distribution tool for marine mammals developed by Kaschner et al., 2006). It was developed for the mass-production of predicted distributional ranges of marine organisms from global occurrence databases, using simple and pre-defined numerical descriptions of species–habitat relationships that can be modified where needed. Predictions from AquaMaps for 12 selected marine fish and mammal species are compared alongside those generated from a range of other methods (GARP, GLM, GAM, MAX-ENT) that are commonly used to construct species distribution models but which are limited in the extent to which experts can influence model parameterisation. Model comparisons were made using independent data from fisheries trawl surveys conducted in UK and Australian waters and dedicated marine mammal surveys in Antarctic waters and in the North Sea.

The objective of the assessment was to compare the performance in terms of predictive accuracy of AquaMaps, a system that can be automated to a great extent and allows the speedy processing of large number of species, with a range of popular and generally more sophisticated routines. If, at the scale of entire species ranges, AquaMaps can produce similarly reliable and verifiable results as commonly used high-end methods, then its greater transparency, ability to incorporate expert knowledge and its online accessibility would facilitate the broad application of such an approach, increasing practical use in the context of decision making and planning processes.

2. Materials and methods

2.1. Marine species occurrence data

Global occurrence data for model building were obtained from two sources. For marine fish, occurrence records were extracted from FishBase, the most comprehensive, online database on fish occurrence records from museum collections and selected, regional trawl surveys (Froese and Pauly, 2007). Marine mammal occurrence records were obtained from OBIS (OBIS, 2007). Similar to FishBase, OBIS is a comprehensive, online database of occurrence data from national museum collections and other sources.

For the marine fish, the species selected represented a broad range of taxa and life histories and were species which were also relatively well represented in the two regions used for model testing, i.e. UK and Australian waters (Table 1). Nine fish species were selected: four that were adequately represented in fisheries surveys conducted in UK waters by the Centre for Environment, Fisheries, and Aquaculture Science (Cefas); four that were adequately represented in fisheries surveys conducted in Australian waters by the Commonwealth Scientific and Industrial Research Organisation (CSIRO); and one (John dory, *Zeus faber*) that was represented in both regions. Raw occurrence data (all accumulated occurrence data per species) from FishBase were extracted for these species. Records deriving from CSIRO surveys were removed, as this data would form the test data for validating the models (Cefas survey data, also used for testing, is not yet represented in FishBase or OBIS and so did not need removing). Occurrence records were spatially aggregated at a resolution of 0.5° latitude × 0.5° longitude and assigned a unique *c*-squares code (Rees, 2003). These could then be converted to a binary format that distinguishes between presence and absence in each cell as input for most subsequent analyses. The exception to this is the testing of predicted gradients of species occurrence with independent survey data where proportional data is used. *c*-squares is a global, spatial indexing system that allows geographic features to be referenced at multiple spatial resolutions, and provides the framework for the database structure behind AquaMaps. Using a fixed spatial resolution and indexing system facilitated the process of constructing and testing the models as data could easily be passed between the various programs containing the modelling routines (see below). Having assigned raw occurrence records to 0.5° *c*-squares cell, potentially erroneous cells were removed if they were: (i) located entirely over land; or (ii) located outside of UN Food and Agriculture Organisation (FAO) fisheries reporting areas where the species is known to occur; or (iii) located outside of expert defined geographic range extents (bounding boxes). FAO areas and bounding boxes were assigned to species using information on species distributions from the many references listed in FishBase (for fish) and those provided in Kaschner et al. (2006), Appendix 2 (for marine mammals). This process is automated in AquaMaps. Further cleaning of data to check for other errors in digitisation, misidentification or data corruption requires significant human input. As the ability of different modelling methods is to be assessed based on their capacity to deal with publicly available data with maximal automation to produce reasonable predictions, such further cleaning was not performed for training data. Test data from surveys are assumed to have minimal error as they came direct from the data source, though basic tests for error in digitalisation were performed. Certain types of error, such as misidentification, will remain in almost any dataset not prepared entirely by a taxonomic expert from original samples.

Three marine mammal species were selected for model comparison (Table 1). These species were chosen due to the contrasting geographic ranges they are known to occupy and the availability of sufficient occurrence data needed for model constructing and testing. Records were treated similarly to those for marine fish.

Table 1

Marine fish and mammal species occurrence data used for model training and survey data used for model testing. Common name taken from FishBase (for fishes) and OBIS (for mammals).

Species code	Species name	Common name	Category	Characteristic description	Model training data (# presence cells)	Test data region	Model testing data		
							Presences (# cells)	Absences (# cells)	Prevalence
HYPLA	<i>Hyperoodon planifrons</i>	Southern bottlenose whale	Mammal	Beaked whale	37	Southern Oceans	468	12,425	0.04
CAEQU	<i>Carangoides equula</i>	Whitfin trevally	Fish	Benthopelagic	40	Australia	49	246	0.17
PSEUR	<i>Psettodes erumei</i>	Indian spiny turbot	Fish	Flatfish	100	Australia	87	248	0.26
CLHAR	<i>Clupea harengus</i>	Herring	Fish	Small pelagic	119	UK	213	353	0.38
SASAG	<i>Sardinops sagax</i>	South American pilchard	Fish	Small pelagic	128	Australia	32	80	0.29
SOSOL	<i>Solea solea</i>	Common sole	Fish	Flatfish	140	UK	104	115	0.47
TRTRA	<i>Trachurus trachurus</i>	Horse mackerel	Fish	Benthopelagic	157	UK	294	353	0.45
SQMEG	<i>Squalus megalops</i>	Shortnose spurdog	Fish	Elasmobranch	216	Australia	70	260	0.21
ZEFAB	<i>Zeus faber</i>	John dory	Fish	Benthopelagic	502	Australia and UK	153	565	0.21
PHPHO	<i>Phocoena phocoena</i>	Harbour porpoise	Mammal	Porpoise	509	North East Atlantic	177	457	0.28
SQACA	<i>Squalus acanthias</i>	Piked dogfish	Fish	Elasmobranch	1468	UK	219	353	0.38
BAPHY	<i>Balaenoptera physalus</i>	Fin whale	Mammal	Baleen whale	1949	Southern Oceans	102	12,425	0.01

2.2. Environmental data

The following global coverage environmental datasets were prepared at 0.5° resolution (259,200 cells). The intention was to agglomerate global maps of a number of key environmental variables based on long-term average conditions using comprehensive, publicly accessible raster data. All geospatial data manipulation and analysis was performed using ArcMap v.9 (Environmental Systems Research Institute). Data is generally available at greater resolutions than the 0.5° resolution used here, and was converted to such by calculating mean, minimum and maximum values, and used as appropriate for mean, minimum and maximum layers.

- Maximum, minimum, and mean depths. Data were extracted from the global coverage ETOPO2 2 min resolution bathymetry dataset (NOAA, 2006).
- Mean annual sea surface temperature (SST) in degrees Celsius covering the period 1982–1999. Data were extracted from a climatology produced following the methods described by Reynolds and Smith (1995) and published by NOAA (2007).
- Mean annual salinity covering the period 1982–1999. Data were extracted from the 2001 World Ocean Atlas (Conkright et al., 2002) published by NOAA.
- Mean annual proportional ice cover (by area) on a scale of 0.00–1.00 and covering the period 1990–1999. Data were obtained from the U.S. National Snow and Ice Data Centre (Cavalieri et al., 2006). Inverse distance weighted interpolation was performed to fill missing data values in a small number of coastal cells (approximately 1000 cells).
- Mean annual primary production in mgC m⁻² day⁻¹ for the period 1997–2004. Data were obtained from the European Joint Research Council (<http://marine.jrc.ec.europa.eu/made> available by Frédéric Mélin) having been generated from remotely sensed chlorophyll *a* concentrations using an approach described in Carr et al. (2006).

The environmental datasets and metadata are freely available at <http://www.aquamaps.org>. Data sets were chosen for their applicability at the global scale and likely variability at the resolution used. Slope and terrain variability measures were considered, but not used. This is because their value for modelling pelagic species may not be good, their variability within cells at the 0.5° resolution may be very great, and there was a desire to maintain clarity and transparency by using only a moderate number of layers. Max-

imum and minimum values for SST, salinity, proportional ice cover and primary production are also available but represent temporal variation within the cell rather than physical variation (as in the case of depth). For simplicity and transparency they have not been included.

2.3. Test data

To test the models we used independent data on presences, absences, and relative abundance collected from targeted surveys. For the marine fish species, data were provided from two regional surveys covering UK and Australian waters. In the UK, data from 5 annual trawl surveys were extracted from the Cefas trawl database (CEFAS, 2007). Catch data from Australian waters were provided from the CSIRO trawl database (MarLIN, 2007). As all of the surveys used different trawl gear, we only used catch data from surveys where the species were susceptible to the gear. Catch data were converted to presences and absences within 0.5° c-squares and also represented as average annual catch rates (kg hr⁻¹ trawl time) at the same resolution. Marine mammal survey data were obtained from two sources: the SCANS survey for the harbour porpoise *Phocoena phocoena* (Hammond et al., 2002), and the International Whaling Commission IDCR-DESS SOWER survey (IWC, 2001) for the Southern Bottlenose Whale *Hyperoodon planifrons* and the Fin Whale *Balaenoptera physalus*. Survey data were processed as described in Kaschner et al. (2006) to compute 'Sightings per unit effort' (SPUE).

2.4. Model construction

We compared AquaMaps with some of the most common methods for generating species predictions models: the Genetic Algorithm Rule-set Procedure (GARP), maximum entropy (Maxent), generalised linear modelling, and generalised additive modelling. A list of methods used and source software is given in Table 2.

2.4.1. AquaMaps

AquaMaps is an automated and adapted version of the Relative Environmental Suitability (RES) modelling approach of Kaschner et al. (2006) which was specifically developed to deal with the data paucity that currently precludes the generation of large scale species distribution for almost all marine mammal species. Predictions of the natural occurrence of a species are gener-

Table 2
Modelling methods used in this study.

Code	Method	Software	Source
AMG	'non-expert AquaMaps'	AquaMaps desktop version	Copy available from lead author
AMEG	'expert AquaMaps'	AquaMaps desktop version	Copy available from lead author
GAM	Generalised Additive Modelling (with random absences)	R statistical software	http://cran.r-project.org/
GLM	Generalised Linear Modelling (with random absences)	R statistical software	http://cran.r-project.org/
MAX	Maxent	Maxent	http://www.cs.princeton.edu/~schapire/maxent/
OMG	GARP best subsets (new implementation)	openModeller	http://openmodeller.sourceforge.net/

ated from pre-defined 'environmental envelopes' that numerically describe a species response to an environmental gradient based on published information about species-specific habitat usage. The combination of these responses in every cell determines the suitability of that cell for the species. Similarly to the RES modelling approach AquaMaps relies on a pre-defined, trapezoidal shape (Fig. 1 here and Figs. 2 and 5 of Kaschner et al., 2006) to describe the basic relationship between species occurrence by means of a preferred range and an absolute range representing the limits of tolerance with respect to a set of equally pre-defined environmental predictors. In addition, AquaMaps does not go through an iterative model selection process which allows for non-linear complex interactions between different predictors, but computes overall probabilities using a simple generic multiplicative model (see below). Hardwiring of the shape of environmental envelopes, predictor selection and model definition was used to (a) maximize transparency and facilitate intuitive understanding of species response curves and predictor interactions for non-modellers (a pre-requisite for expert review and identification of sampling biases), (b) speed up computational processing, and (c) maintain clarity in the reproducibility of the results (using the same envelopes, algorithm and environmental data, any GIS or database system should be able to produce the same output).

Anchor points for the species-specific absolute and preferred habitat ranges (i.e. environmental envelopes) are calculated based

on a subset of available presence cells that have been subjected to a series of location-based quality checks (see above). The environmental envelopes are computed from the environmental values of the locations at which the species is found to occur using the following rules:

- Absolute minimum (MinA) = the 25th percentile of the environmental values – (1.5 × the interquartile range), OR the absolute minimum environmental value at which the species is observed, whichever is lower
- Preferred minimum (MinP) = the 10th percentile of the environmental values
- Preferred maximum (MaxP) = the 90th percentile of the environmental values
- Absolute maximum (MaxA) = the 75th percentile of the environmental values + (1.5 × the interquartile range), OR the absolute maximum environmental value at which the species is observed, whichever is greater

Computed environmental envelopes for each species can be viewed alongside their maps online through <http://www.aquamaps.org> with some examples shown in Fig. 1, while values for maxima and minima for all species in this analysis are described in Table 3. After the definition of environmental envelopes, predictions of species-specific relative habitat suitability are generated for each 0.5° grid cell by assigning a probability of '1' to all cells asso-

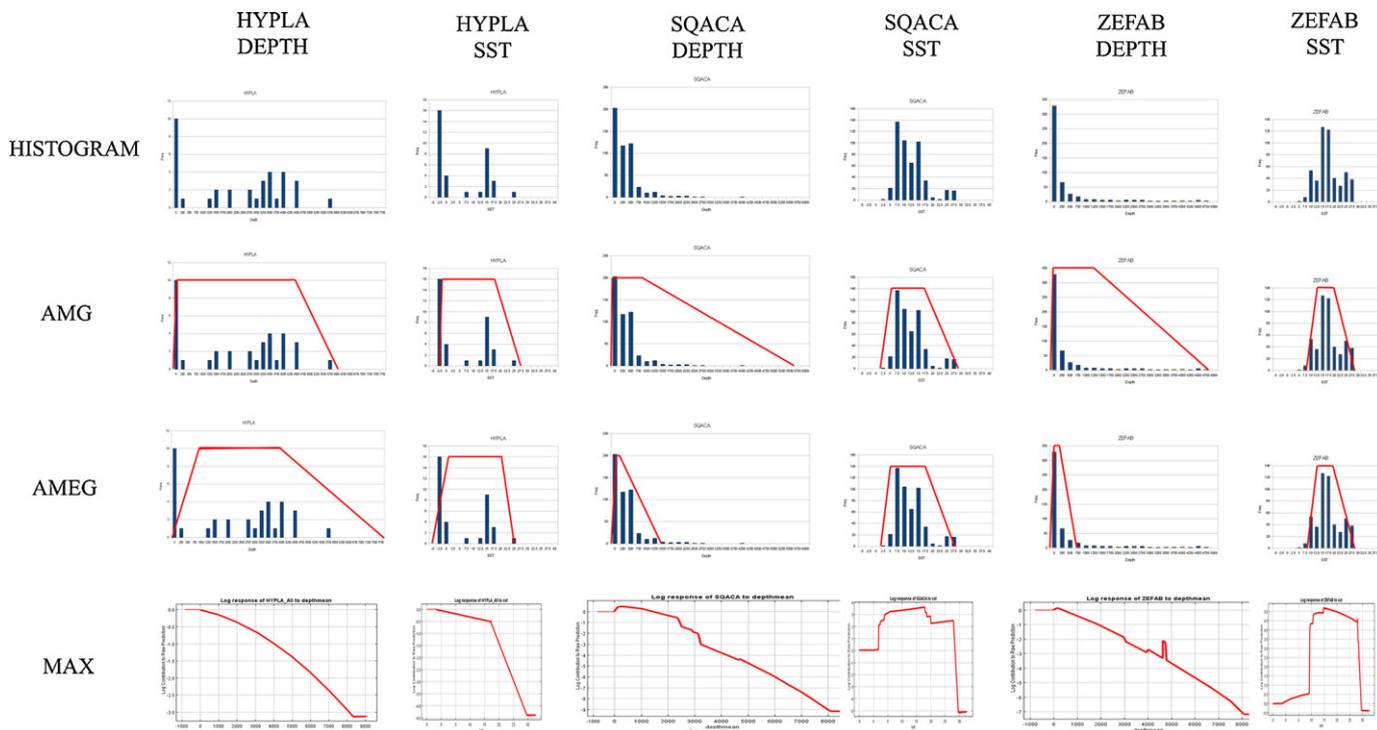


Fig. 1. Comparison of the envelopes of AquaMaps (AMG and AMEG) with the response curves generated by Maxent. Histograms represent frequencies of presence points for classes of the environmental variable. Envelopes or response curves for depth and sea surface temperature are shown for one mammal species (HYPLA) and two fish species (SQACA and ZEFAB).

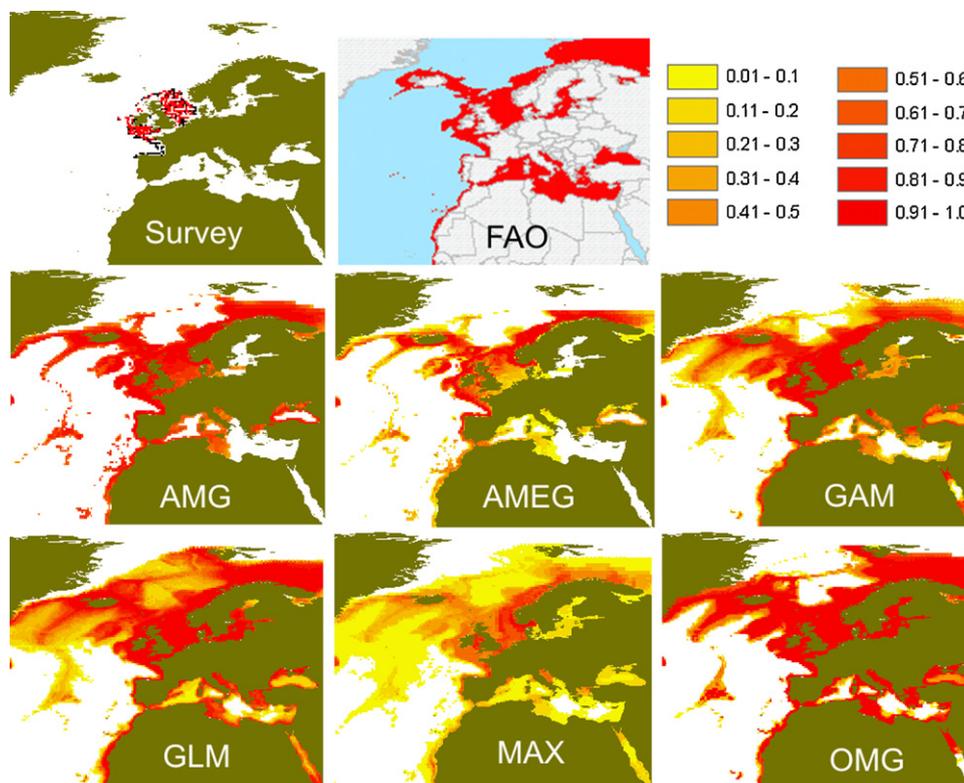


Fig. 2. External test data survey locations (presences in red), current FAO species actual range map (Compagno, 1984), class intervals of likelihood of occurrence in modelled distributions, and modelled distributions for *Squalus acanthias* in the NE Atlantic. Models labelled as in Table 2.

ciated with environmental values falling between the preferred minimum and preferred maximum of a specific envelope, and '0' to all values \leq absolute minimum and \geq absolute maximum. Environmental values falling in the range between absolute minimum to preferred minimum and preferred maximum to absolute maximum are assigned a probability of >0 and <1 using linear interpolation. Overall species-specific relative habitat suitability for each grid cell is then computed as the geometric mean of all individual predictor probabilities, assuming an equal weighting for all.

We generated predicted distributions for the 9 fishes and 3 mammals from environmental envelopes that were exclusively based on available occurrence records alone and did not incorporate additional information such as known depth ranges or other types of expert knowledge. We referred to this method as 'non-expert AquaMaps' (AMG – Table 2).

In the online version of the AquaMaps approach, additional rules are incorporated to automatically modify the minimum and maximum preferred and absolute ranges for a species based on an understanding of its ecology and known distribution. Pelagic fish species are modelled so that both preferred maximum and absolute maximum depths are assigned the maximum depth of the world's oceans, as increasing depth most likely will not affect their distributions. Further, expert information such as known published depth ranges for species can be automatically imported from existing databases (as for all fishes in this analysis – data from FishBase, Froese and Pauly, 2007), or can be manually entered into the AquaMaps system from literature sources (as for mammals in this analysis – data from Kaschner et al., 2006). Depth envelopes are then calculated using minimum and maximum depth values for each 0.5° cell rather than using a single layer (mean depth). A species' tolerance with respect to preferred minimum and absolute minimum depths are related to the environment's maximum depth attribute in

each cell while the species' tolerances with respect to preferred maximum and absolute maximum depths are related to the environment's minimum depth attribute in each cell. In addition to these default incorporations of published habitat use information, individual environmental envelopes can be modified manually if published information suggests that existing point occurrences do not provide a representative coverage of a given species distribution in environmental space. We refer to the method of producing species distribution maps using such additional information as 'expert AquaMaps' (AMEG – Table 2). Predictions were also made for the 9 fish and 3 mammal species using this method. The modifications made for each species are given in Table 3.

All permutations of the AquaMaps methodology were implemented in a desktop version of the software, copies of which are available from the lead author. Note that the online version of AquaMaps (<http://www.aquamaps.org>) is only available in the form of the basic 'expert AquaMaps' method which incorporates independent information about depth preferences and maximum latitudinal and longitudinal boundaries that are available from FishBase. Envelope settings for all other environmental parameters have only been reviewed on a case-by-case basis. Given the same input settings to AquaMaps, the outputs will always be the same. This in contrast to some other methods such as GARP, which produce slightly different outputs with the same input information.

2.4.2. GARP (genetic algorithm for rule-set production)

GARP has received considerable attention by ecological modellers, and has been used to model the distributions of a wide range of terrestrial and marine organisms. Here, we ran GARP in 'best subsets' mode (Anderson et al., 2003) as implemented in the openModeller package (Muñoz et al., 2007). Settings were mainly

Table 3
Environmental values defining the environmental envelopes used by AquaMaps. Depth use is different under 'expert AquaMaps' (using maximum and minimum depth in each cell instead of mean depth). Expert modified values are highlighted and their original 'non-expert AquaMaps' values are shown in parentheses. * = changes from the values reported in Kaschner et al. (2006), na = not used for predicting mammal distributions.

Species code	Depth (m)		Mean SST (°C)				Mean salinity				Mean ice cover (proportion of cell)				Mean annual primary prod (mg C m ⁻² day ⁻¹)			
	MinA	MinP	MinA	MinP	MaxA	MaxP	MinA	MinP	MaxA	MaxP	MinA	MinP	MaxA	MaxP	MinA	MinP	MaxA	MaxP
CAEQU	100 (9)	110 (40)	17.50	22.42	29.02	28.52	33.38	34.00	37.00	35.38	0	0	0	0	48	424	1190	2409
TRTRA	0 (1)	100 (18)		12.00	28.42	21.94	6.46	33.36	37.75	39.49	0	0	0.02	0.05	0	425	1891	3338
PSERU	1 (3)	20 (12)	0.56	9.65	29.49	28.81	29.77	32.81	35.16	38.66	0	0	0	0	87	406	1460	2111
SOSOL	0 (1)	10 (11)	6.97	9.76	25.00	18.98	7.59	33.43	37.81	39.09	0	0	0.03	0.13	0	512	1859	2628
SQACA	0 (1)	192 (37)	1.01	5.22	27.25	17.36	7.59	31.30	35.09	38.25	0	0	0.05	0.44	100	417	1667	3988
SOMEG	30 (1)	118 (41)	7.70	15.22	29.22	25.67	34.32	34.90	35.90	38.11	0	0	0	0	69	364	1418	3392
CLHAR	0 (1)	22 (6)	-1.30	-0.10	24.70	9.34	3.55	6.77	35.06	36.13	0	0	0.22	0.44	464	630	1862	2863
SASAG	0 (2)	22 (18)	4.62	10.41	28.73	23.92	30.27	32.11	35.65	37.92	0	0	0	0.20	136	484	1860	2854
ZEFAB	5 (1)	50 (20)	9.60	12.00	28.02	19.81	19.62	33.00	36.08	39.01	0	0	0	0.03	0	432	1669	3830
BAPHY	0 (1)	1000 (95)	-5*	0	25	20	na	na	na	na	0	0	0.50	na	na	na	na	na
			(-1.79)	(-0.71)	(28.93)	(13.22)	na	na	na	na	0	0.01 (0)	(0.37)	(0.94)	na	na	na	na
HYPLA	0 (12)	1000 (43)	-5*	0	25	20	na	na	na	na	0	0.01 (0)	0.50	0.90	na	na	na	na
			(-1.65)	(-1.43)	(27.37)	(18.01)	na	na	na	na	0	0	(0.49)	(0.62)	na	na	na	na
PHPHO	0 (1)	10 (22)	5 (4.28)	5 (4.28)	20 (24.33)	15	na	na	na	na	0	0	0.01	0.50	na	na	na	na
			(-1.74)			(12.79)							(0.13)	(0.73)				

left as default, i.e., a training proportion of 0.5, 2500 re-samples, a population size of 50, 20 models under omission, a hard omission threshold of 100, a convergence limit of 0.01, a commission threshold of 50, and a commission sample of 10,000. The number of runs was increased to 100. Greater differentiation in prediction values is possible by increasing run number, but computing times increase exponentially and there is reportedly little gain in model quality above 100 runs (Siqueira and Durigan, 2007). When 100 runs are performed, models took over five times longer to produce than for all other methods. Exact times varied depending on the species being modelled.

2.4.3. Maxent (maximum entropy modelling software)

Maxent was developed by Phillips et al. (2006) and is based on the principles of maximum entropy, whereby a target probability distribution is estimated by finding the probability distribution of maximum entropy, i.e., that is most spread out or closest to uniform, subject to a set of constraints that represent incomplete information about the target distribution. Models were constructed using the Maxent software (Table 2) set to the following default parameters: use of a random seed, minimised memory use, a random test percentage of 0, a regularization parameter of 1, a maximum of 500 iterations, a convergence threshold of 0.00001 and a maximum of 10,000 background points, with no test sample or bias file.

2.4.4. GLM (Generalised linear models) and GAM (Generalised additive models)

Generalised linear (GLM) and additive models (GAM) (McCullagh and Nelder, 1989; Hastie, 1991) were constructed using the open-source statistical programming language R version 2.5.1 (<http://cran.r-project.org/>). To estimate GLMs and GAMs, absence data were first generated by selecting 0.5° cells at random to arrive at 50% prevalence for each species. GLMs were fitted to the presence-absence data using a logistic link function. GAMs were also generated using a logistic link function with cubic splines applied to all predictors. To allow a consistent application across species, as in other modelling approaches, models were applied naively, without interactions, and to the full set of predictors without prior testing for correlation.

2.5. Model testing

Models were tested in two ways. For the first test, estimates of ROC-AUC were generated between the model predictions and presence/absence test data using the ROC package for R (Sing et al., 2005). ROC is the receiver operating characteristics curve, with AUC being the area under curve value (Swets, 1988). Construction of distribution models for a large number of marine species, many of which may not be well represented in targeted, independent surveys means that in most cases prediction accuracy can only be assessed using an internal test. The ROC-AUC test was therefore run using predictions generated from subsets of the original occurrence data and tested with the remaining occurrence data. For each fish and mammal species we generated four subsets of occurrence data each comprising of 75% of data points chosen via random selection. Prediction accuracy was assessed by comparing the remaining 25% of data points with the predictions generated from its counterpart subset. This is a cross-validation ROC (Fielding and Bell, 1997). This variation of the ROC-AUC test could not be used to test AMEG predictions as the expert review process would act to standardise the environmental envelopes generated from the four different subsets. The ROC-AUC test was also run using model predictions generated from the full occurrence datasets and tested with independent test data, with the latter summarised as presence or absence per 0.5° cell. This test could also be run on all modelling methods. Since ROC-AUC has been shown to be largely unaffected

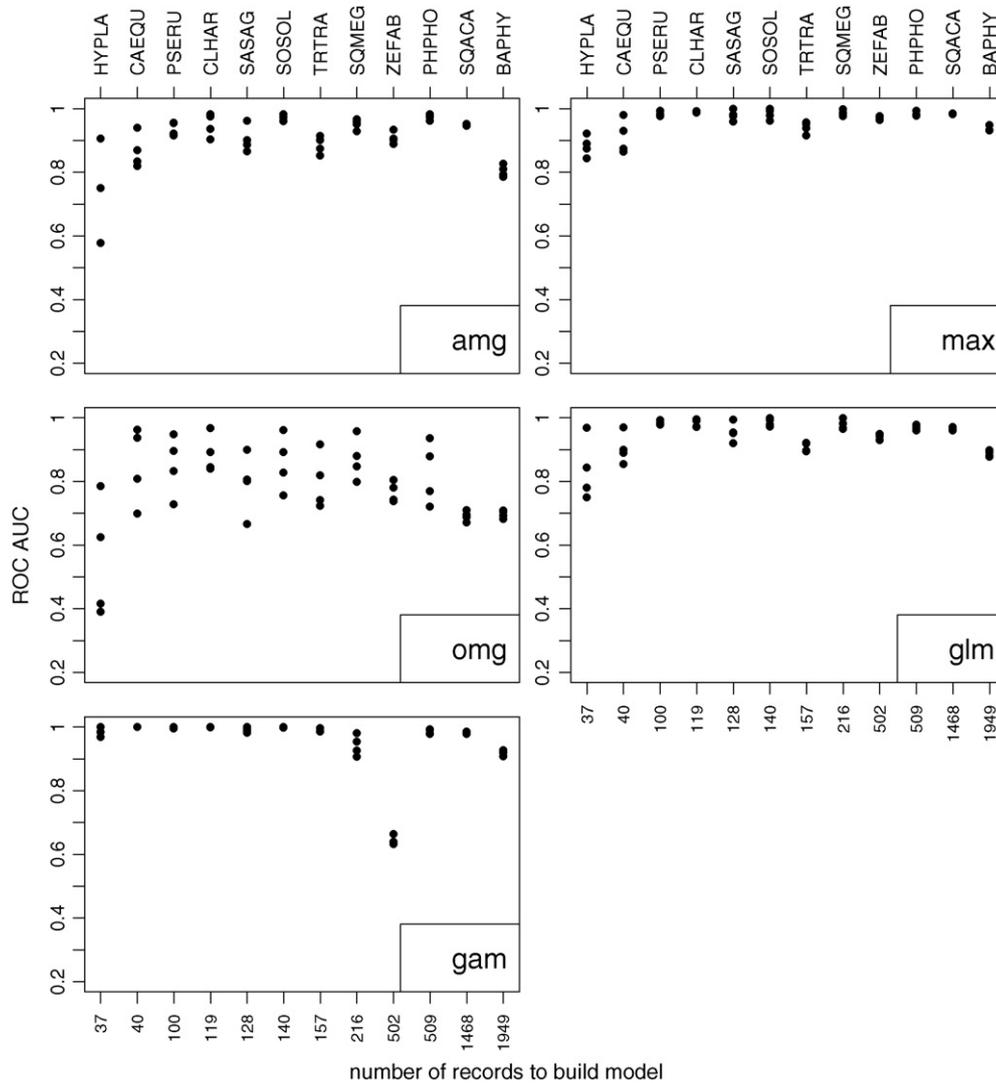


Fig. 3. ROC–AUC internal test results show variation between modelling methods across the range of input data quantity and species. Model predictions are based on four different 75% subsets of the original occurrence data for each species. No test was performed for AMEG as expert adaption results in a single output. Species are labelled as in Table 1 and ordered left to right based on increasing training data sample size.

by species prevalence, it allows a direct comparisons of models for different species (McPherson et al., 2004).

The second test was used to assess to what extent predictions of relative habitat suitability produced by the different models matched observed indices of effort-corrected species occurrence. i.e. to compare gradients of predicted and observed species occurrence. The test we used was developed by Kaschner et al. (2006) following an approach recommended by Pearce and Boyce (2006) to test predictions of presence-only models. Spearman’s rank correlations are computed between model predictions and relative abundance/density estimates based on the average effort-corrected catch or sighting rates over all cells within classes of predicted probability. To assess the performance of our models compared to random distributions, we obtained a simulated *p*-value by recording the number of times the relationship between 1000 random data sets and test data sets was as strong as or stronger than that found between the observed encounter rates and the model predictions.

3. Results

The various model algorithms generated different predictions for each species, and the envelopes and response curves for some

of these are presented for comparison in Fig. 1. Fig. 2 shows the model outputs for *Squalus acanthias* along with the survey data collection locations used for external testing of the models and the current FAO distribution map for the same species in the same area (Compagno, 1984). In all cases the models predicted *S. acanthias* to occur far beyond the geographic range described by the input data. The AMG model describes a similar pattern to the known distribution, though with some restriction of range in areas of extremely high and low salinity (Mediterranean, Red and Baltic seas). The AMEG model is similar in overall extent (area) but shows a significant restriction in probabilities with depth. The difference in this respect is clearly seen in the expert modifications for depth across all species (Table 3 and Fig. 1). The GAM model is quite similar to the AMG model, though indicates a greater proportion of less suitable environments. The GLM and MAX models cover a greater range than the other models, with the difference between them being that the GLM model predicts a relatively large proportion of the range as having a high probability of occurrence while the MAX model predicts only a small proportion of the overall range will have a high probability of occurrence. The OMG model predicts much the same range as the GAM model, though it predicts nearly all of the range to have a high likelihood of occurrence, with variation in likelihood of occurrence being restricted to the peripheral areas of the

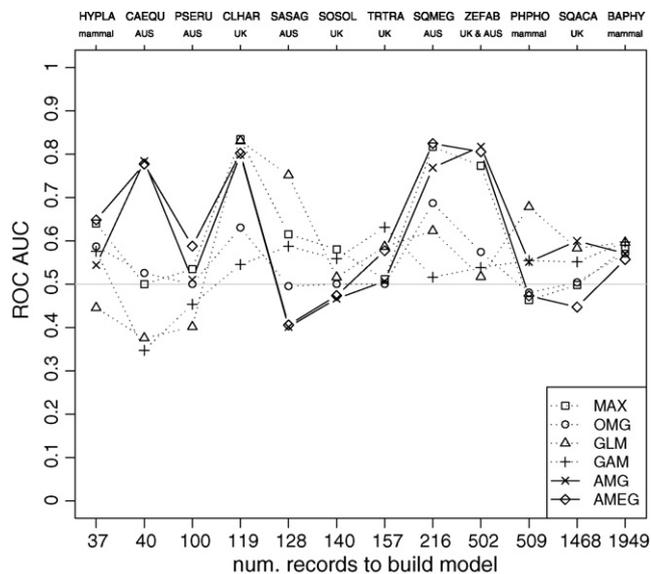


Fig. 4. ROC–AUC statistical results for model predictions based on all points tested against external survey data. Species are labelled as in Table 1 and ordered left to right based on increasing training data sample size.

range. The variation between model predictions in terms of range and the proportion of the range predicted to have a high likelihood of occurrence is generally similar for the other species. The exceptions are AMEG models where, apart from similar changes based on use of maximum and minimum depths (Table 3), variation is more dependent on the degree of changes due to expert modified envelopes.

Internal testing of predictions using ROC–AUC scores for four subsets testing against remaining non-independent occurrence data (Fig. 3) showed that modelling methods generally provided better outputs for species with more data (as seen for some models where results are poor using the lowest numbers of training data). Deviations from this trend with large amounts of training data indicate points of note. ROC–AUC scores are lower and more variable with subsets of 37/40 occurrence points and generally high and not as variable with subsets of 100+ occurrence points. The clear exception is for *B. physalus* which generally performed slightly less well than might be expected for all models given the large amount of training data. *Trachurus trachurus*, *Sardinops sagax* and *Z. faber* also showed scores that were marginally lower than the general trend. When comparing different models, the GAMs obtain remarkably high scores with little variation. Exceptions to this are for the taxa *Squalus megalops* and *Z. faber*, where the distinct drop in ROC–AUC scores is consistent with a limited number of classes (2) generated by the GAM model predicted distribution. The low number of classes of outputs also explains the non-valid results of Spearman's rank test for these taxa when this model is compared with the species survey abundance estimates (see Section 4).

External testing through calculation of ROC–AUC scores from comparison of predictions with external independent test data (Fig. 4) are very variable and generally not very high, even when training data quantity is maximised. Some species appear to be modelled well at least by some modelling methods (comparatively better ROC–AUC values). Comparatively better ROC–AUC scores (>0.7) were obtained for *Carangoides equula*, *Clupea harengus*, *S. megalops* and *Z. faber* models produced by both AMG and AMEG while MAX models of *C. harengus*, *S. megalops* and *Z. faber* also scored comparatively well. GLMs produced comparatively good models for *C. harengus* and *S. sagax*, while GAM and OMG methods produced no models gaining such ROC–AUC scores.

External testing comparing relative predicted probabilities with relative abundances of species from survey data varied in a similar way to ROC–AUC scores in most respects with good correlations possible when models are made from as little as 37 occurrence points (Table 4 and Fig. 5). Overall, models tended to perform better in terms of predicting gradients of relative species occurrence as indicated by the higher number of statistically significant correlations with external survey data. Interestingly, GAMs, which performed extremely poorly in terms of predicting binary presence/absence of species, was one of the methods able to predict relative species occurrence most reliably. OMG models again performed poorly. Again, the low number of output classes produced from these models made some comparisons (four species) statistically impossible. It should be noted that there was no significant result for OMG models of *C. harengus* and *B. physalus*, which were modelled well by almost all other methods.

Expert review effects on predictions based on results in Table 4 and Figs. 4 and 5 show that for most species for which AMG models were good predictors, the main effect in 'expert AquaMaps' was an improvement in ROC–AUC score or correlation with survey data, and that this is largely related to the incorporation of expert defined depth preferences and the different use of depth data in AMEG where a species' response to depth can be applied to both maximum and minimum depths of a cell instead of just mean depth of the cell (Table 3).

A summary of the results of both ROC–AUC and Spearman's rank statistics gives model/species combinations which have both a good range as determined by the ROC–AUC score and a good correlation with existing abundances (Table 5). Comparing model method results, AMG predictions vary to some extent with species, but produce reasonable results under both testing methods and AMEG predictions performed generally slightly better than AMG. GAM predictions performed very poorly at prediction of presence vs. absence (ROC–AUC), but very well at producing predictions which correlate with survey abundances. GLM predictions showed a similar trend but performed less successfully overall. Maxent models performed successfully for a similar number of species as AMG though for a slightly different set of species, and OMG predictions performed poorly.

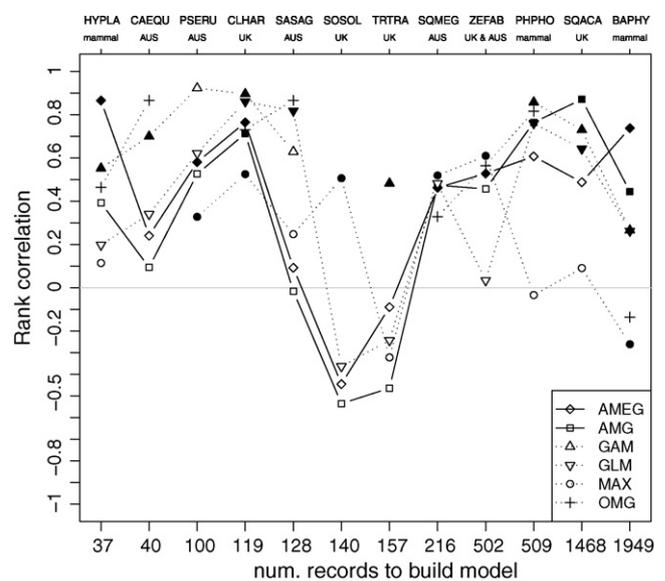


Fig. 5. Results of Spearman's rank correlation of predicted models with survey abundances (significant correlations highlighted in Table 4 have filled symbols). Species are labelled as in Table 1 and ordered left to right based on increasing training data sample size.

Table 4

Spearman's rank correlation (Rho) between prediction models and survey calculated abundances. *P* (Rho) is the probability value for the correlation and *P* (Boot) is the probability that the strength of the correlation between model probabilities and survey abundance is significantly different from the chance of correlation of 1000 randomly generated datasets with the survey abundance values. Significant positive correlations are those where Rho is positive and both *P* (Rho) and *P* (boot) are both less than 0.05 (emboldened in table).

Species	HYPLA	CAEQU	PSERU	CLHAR	SASAG	SOSOL	TRTRA	SQMEG	ZEFAB	PHPHO	SQACA	BAPHY	Number of significant correlations by model
Training data (# Presence cells)	37	40	100	119	128	140	157	216	502	509	1468	1949	
Location	SO	Aust	Aus	UK	Aus	UK	UK	Aus	Aus + UK	NEAtl	UK	SO	
Rho-AMG	0.392	0.094	0.526	0.712	-0.017	-0.536	-0.464	0.473	0.456	0.764	0.870	0.444	
<i>P</i> (Rho)-AMG	0.001	0.592	0.119	0.000	0.957	0.236	0.302	0.009	0.013	0.027	0.002	0.000	4
<i>P</i> (Boot)-AMG	0.186	0.725	0.182	0.000	0.953	0.980	0.987	0.007	0.123	0.218	0.046	0.008	
Rho-AMEG	0.865	0.240	0.580	0.764	0.092	-0.445	-0.090	0.462	0.528	0.607	0.487	0.738	
<i>P</i> (Rho)-AMEG	0.000	0.185	0.001	0.000	0.734	0.173	0.636	0.001	0.001	0.028	0.003	0.000	6
<i>P</i> (Boot)-AMEG	0.000	0.193	0.002	0.000	0.845	0.976	0.875	0.026	0.010	0.299	0.083	0.005	
Rho-GAM	0.552	0.699	0.922	0.895	0.629	N/A	0.482	N/A	N/A	0.857	0.730	0.266	
<i>P</i> (Rho)-GAM	0.000	0.000	0.001	0.000	0.012	N/A	0.000	N/A	N/A	0.000	0.000	0.008	7
<i>P</i> (Boot)-GAM	0.000	0.007	0.054	0.000	0.166	N/A	0.009	N/A	N/A	0.001	0.044	0.003	
Rho - GLM	0.198	0.340	0.621	0.860	0.817	-0.364	-0.243	0.482	0.034	0.759	0.642	0.262	
<i>P</i> (Rho)-GLM	0.064	0.198	0.100	0.000	0.000	0.273	0.152	0.002	0.842	0.000	0.000	0.010	5
<i>P</i> (Boot)-GLM	0.143	0.686	0.466	0.000	0.012	0.966	1.000	0.392	0.952	0.019	0.027	0.000	
Rho-MAX	0.113	N/A	0.327	0.524	0.247	0.506	-0.323	0.518	0.609	-0.035	0.090	-0.262	
<i>P</i> (Rho)-MAX	0.317	N/A	0.003	0.000	0.135	0.000	0.010	0.000	0.000	0.747	0.493	0.009	5
<i>P</i> (Boot)-MAX	0.024	N/A	0.024	0.000	0.148	0.000	0.705	0.000	0.000	0.721	0.212	0.020	
Rho-OMG	0.464	0.866	N/A	0.730	0.866	N/A	N/A	0.329	0.564	0.815	N/A	-0.136	
<i>P</i> (Rho)-OMG	0.150	0.333	N/A	0.025	0.333	N/A	N/A	0.388	0.322	0.025	N/A	0.694	0
<i>P</i> (Boot)-OMG	0.198	0.134	N/A	0.083	0.143	N/A	N/A	0.456	0.085	0.104	N/A	0.695	

Table 5

Summary of species which were modelled reasonably by each algorithm as determined by both ROC-AUC and Spearman's rank statistics (external test of ROC-AUC >0.7, Spearman's Rank positive correlation and significant as highlighted in Table 4) * indicates species with known issues regarding taxonomic uncertainty of different populations or variable catch success in surveys due to different substrates and the effectiveness of gear at different depths. Yes/Possible indicates the fraction of models produced by each method which performed reasonably under each test.

Species	AMG		AMEG		GAM		GLM		MAX		OMG	
	ROC-AUC	Rho	ROC-AUC	Rho								
HYPLA	No	No	No	Yes	No	Yes	No	No	No	No	No	No
CAEQU	Yes	No	Yes	No	No	Yes	No	No	No	No	N/A	No
PSERU	No	No	No	Yes	No	NO	No	No	No	No	Yes	No
CLHAR	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No
SASAG*	No	No	No	No	No	No	Yes	Yes	No	No	No	No
SOSOL*	No	No	No	No	No	No	N/A	No	No	No	Yes	No
TRTRA*	No	No	No	No	No	Yes	No	No	No	No	No	N/A
SQMEG	Yes	Yes	Yes	Yes	No	N/A	No	No	Yes	Yes	No	No
ZEFAB	Yes	No	Yes	Yes	No	N/A	No	No	Yes	Yes	Yes	No
PHPHO	No	No	No	No	No	Yes	No	Yes	No	No	No	No
SQACA	No	Yes	No	No	No	Yes	No	Yes	No	No	No	N/A
BAPHY	No	Yes	No	Yes	No	Yes	No	Yes	No	No	No	No
Yes/Possible	4/12	4/12	4/12	6/12	0/12	7/10	2/12	5/12	3/12	5/11	0/12	0/8
Yes/Possible excluding *	4/9	4/9	4/9	6/9	0/9	6/8	1/9	4/9	3/9	4/8	0/9	0/7

4. Discussion

The aim of developing AquaMaps was to provide a system which can be applied globally and is automated at the basic level, producing distributional range maps for as many species as for which suitable data is available, and from which outputs can be summarised to analyse marine biodiversity in time and space. An easy to understand concept and transparent process that allows expert review of predictions was also an aim, so that poor predictions can be improved by experts on the taxon in question rather than requiring in depth knowledge of the method of modelling. This is typified by the use of environmental envelopes to define the environmental limits of species which is then mapped onto the known environment, rather than the more complex response curves described by many modelling algorithms which define often complex equations relating presence to the predictor variables.

Given these aims, the results of statistical analysis show that AquaMaps compares reasonably (Table 5) with existing niche modelling methods over a range of taxonomic groups, geographical regions and training data sample sizes given the environmental datasets which are easily available for such analyses. Good prediction models can be achieved with the 'expert AquaMaps' method with as little as 37 occurrence points (Table 4), with 'non-expert AquaMaps' providing a reasonable first pass prediction for the range of a species from as little as 40 occurrence points (Fig. 2) with little variance under internal tests (Fig. 3). In terms of the number of points required for modelling, this fits well with the results obtained by Stockwell and Peterson (2002), where between 20 and 50 points were demonstrated to result in reasonable models when using Desktop GARP.

In previous analyses (Elith et al., 2006), Maxent is one of the best performing modelling packages when testing for binary pres-

ence/absence (ROC–AUC). Similar analyses performed here support this result to some extent, though Maxent models were found to perform poorly for some taxa (particularly the mammals – Fig. 3). However, even when performing well under ROC–AUC analysis, the degree of correlation with survey data was not necessarily the highest of the modelling methods used. *C. harengus* was the only species for which almost all modelling methods produced models with statistically significant correlations with survey abundances, and yet the correlation value obtained by Maxent was the lowest (Table 4). The only case in which Maxent was uniquely better than all other models was for *Solea solea* (Tables 4 and 5). This is the only species for which training data came almost exclusively from the same region as the independent survey dataset. This indicates that Maxent may over-fit predictions to the areas where sample occurrence data have been collected. This may be especially true if there is some kind of sampling selection bias to these collection locations (e.g. a preference to catch fish in shallower water of a certain bottom type). Maxent does have an ability to counteract over-fitting using its regularization procedure (Hernandez et al., 2006; Phillips et al., 2006), but it is unclear as to how this might be included in automated mapping of many species. It would only be useful for some species, and it may not be a valuable exercise given assumptions that the sampling is not biased (Phillips, 2008).

GLMs and GAMs are commonly used to model distributions (Guisan and Thuiller, 2005), but rely on input in the analysis to obtain best results. Under the requirements of an automated system, where selection of variables is standardised, these methods appear to perform poorly (Table 5). Neither performs well in standard binary presence/absence tests (ROC–AUC). The GAMs performed relatively well in the tests of correlation with survey data (Tables 4 and 5), although notably for species which other methods could not model well. Of particular note was the correlation of the GAM model for *T. trachurus*. This species may in fact represent more than one species (see discussion below) and as such the production of a model that performs well in such tests may not indicate good model performance.

OMG (GARP) models performed badly in almost all comparisons (Tables 4 and 5). The outputs were generally much more variable than all other methods and only seemed to reduce in variability and obtain a stable ROC–AUC score when over 1000 occurrence points were used to model the distribution, at which point the scores are quite low (Fig. 3). This is likely a result of the lower number of output classes under OMG models. This is seen in Fig. 2 where the predicted area includes a uniform block of red with a high predicted value (0.9–1). It is possible that a higher number of output classes might be producible by increasing the number of runs performed when modelling distributions using OMG. However, this would require modelling time in excess of that available and may not improve the output quality enough to generate the needed classes.

As all models were developed with the same set of environmental layers, variation in statistical support for the models is expected to vary with life history, taxonomic group/status and bias in sampling distribution. This is clearly seen in the statistical analysis. When looking at internal tests with subsets of data (Fig. 2), the slightly lower ROC–AUC score of *Z. faber* is likely due to the fact that it lives over a greater range than all other fish species, and as such sample selection bias may be greater due to more data from better sampled regions. Models for *B. physalus* show a significant drop from the trend based on numbers of occurrence points included, which may in part be due to sample selection bias over a large range for both sampling and surveying options, where both are limited temporally (to different extents) to the Antarctic summer season. Additionally, samples from whaling efforts which are included in the OBIS dataset may be biased to areas known for higher abundances of both this and other species. Modelling performance may

also be poor for specific species if taxonomic uncertainty leads to the inclusion of environments to which different populations (potential taxa) are adapted. This may explain the poor modelling performance for *S. sagax* and *T. trachurus* by almost all modelling systems. *S. sagax* represents a species which has genetically identified sub-populations in different parts of its distribution (Grant et al., 1998), while *T. trachurus* represents a species where misidentifications may occur in part of its range due to presence of a closely related species (southern populations of *T. trachurus* may represent *Trachurus capensis*) (Froese and Pauly, 2007).

S. solea is one of the species that is generally modelled badly when compared to survey data (Table 5), and likely reflects the relative catch of the species being more sensitive to local factors within the survey area itself or the methods used to sample the species. *S. solea* has specific bottom type requirements, especially with regards to depth and sediment type (Rogers, 1992). It is not always caught evenly in surveys as the catch efficiency of the gear, which is dependent on the bottom contact of the trawl and fishing protocols, also varies with depth, bottom substrate and topography. The latter refers to the process by which faster or deeper trawls tend to cause the net to jump up more frequently resulting in poorer catches of bottom dwelling fish. The slightly less 'bottom associated' (Froese and Pauly, 2007) flatfish *Psettoodes erumei* was apparently modelled slightly better with the inclusion of depth ranges in 'expert AquaMaps' resulting in a significant correlation of the model with survey abundances, and the general additive model nearing significant correlation.

The comparison of the expert input in AquaMaps is generally favourable to the inclusion of expert knowledge, but some loss in performance underlines that existing expert knowledge may not always result in improved outputs and supports the sensitivity of habitat rating to expert opinion found in previous analysis (Johnson and Gillingham, 2004). All existing data, including expert knowledge, is prone to bias. Even if it is assumed that all *T. trachurus* occurrence data are valid identifications, bias in occurrence data and/or expert knowledge can explain the poor model performance. Such bias may be sufficiently strong (large survey datasets included from certain regions) to result in poor predictions of suitable habitat in other regions or may also be combined with species-specific phenomena such as tropical submergence (Ekman, 1967) which may also lead to bias in expert knowledge as expert knowledge is generally biased towards surface waters.

The better performance of the 'non-expert AquaMaps' compared to 'expert AquaMaps' for *S. acanthias* and *P. phocoena* is suggested to be a result of a bias in the treatment of depth. For *S. acanthias*, the given expert value for preferred depth range is likely wrong. Textual descriptions of localities where the species is caught indicate a lower value should be applied to preferred minimum depth. Preliminary analysis with such lower values produced predictions more similar to the 'non-expert AquaMaps' model. The result for *S. acanthias* highlights how such analyses can draw attention to possible errors in data presented in the literature and taken as the current state of knowledge. *S. acanthias* was selected for Fig. 2 to highlight the importance of verification of expert knowledge. Alternatively, the need for expert knowledge applied to the Southern Bottlenose Whale, *Hyperoodon planifrons* (HYPLA), is evident in the differences in the depth envelopes of AMG and AMEG (Fig. 1 and Table 3). The species is one of the deep-diving beaked whale species, known to predominantly occur almost entirely in deeper waters (Gowans, 2002; Kasamatsu et al., 2000; MacLeod and D'Amico, 2006). The large number of shallow water sighting of the species reflects known sampling biases of heterogeneous survey efforts in shallow waters and potential misidentifications, as all beaked whales are highly inconspicuous and difficult to identify at sea. In addition, the stranding records, which generally represent the most common form of available occurrence records for most

beaked whale species, would have been allocated to coastal (rather than land) cells and thus would not have been successfully filtered out during the initial screening for erroneous species reports.

The 'expert AquaMaps' envelope for *P. phocoena* had been taken from the previous work of Kaschner et al. (2006) which used average depths and had not been adjusted to the current algorithm's use of minimum and maximum depths. As such the differential use of depth in 'expert AquaMaps' has worsened the correlation between model and survey data (at least within the survey area) while the non-expert mode used occurrence points from roughly the same area as the independent survey dataset to create the model, thus producing a reasonably good approximation. Further expert review given knowledge of the new use of depth values may allow recovery of a stronger, more significant correlation for *P. phocoena*. Both other mammals actually show improvements in statistical support under the new 'expert AquaMaps' prediction when compared to the original RES model of Kaschner et al. (2006), indicating that the method has generally remained effective for the prediction of marine mammal distributions.

If taxa with known issues regarding taxonomic uncertainty or poor catch in survey are excluded from the summary in Table 5, then the 'expert AquaMaps' also compares well with general additive models under tests of correlation with survey abundance.

Lobo et al. (2008) argue that ROC–AUC scores are not a good measure of model performance for five reasons, one of which is that they ignore the predicted probability values and the goodness-of-fit of the model. We use the Spearman's rank statistic to directly test for a good fit between the probability values generated by models and true abundances. The limitation of tests such as this are that they require external survey data for the species studied, including sampling from areas where abundances are lower, to confirm this goodness of fit. Such data are not available for the vast majority of species. Lobo et al. (2008) also state that the total extent to which models are carried out highly influences the rate of well-predicted absences and the AUC scores, with the generation of pseudo-absences for points which are geographically and therefore probably more environmentally distant from the presence localities leading to a low commission error. The statistical comparisons carried out here were based on surveyed regions where the species modelled are known to occur, therefore avoiding this problem. Nevertheless ROC–AUC scores have been widely used in ecological modelling and retain some advantageous features. Lobo et al. (2008) suggest that 'the real value of AUC is that it provides a measure of the degree to which a species is restricted to a part of the variation range of the modelled predictors, so that presences can be told apart from absences' i.e. it tells us whether the range of the predicted distribution is more or less accurate in environmental space, but not whether the internal probabilities are good. Comparison of ROC–AUC results of different models for the same species is possible because the data used to generate and test the predictions remain constant. ROC–AUC also remains useful as a basic statistic to determine the amount of variation in results from partitioned datasets, as performed here, highlighting the potential degree of sampling bias in the original datasets with respect to the environmental parameters used for modelling and the number and distribution of known occurrences used to generate the model.

Expert review remains the quickest way to improve predicted distributions in AquaMaps, but relies on expert knowledge being as complete as possible. Taxonomic uncertainty and poor knowledge of biodiversity in certain geographical areas remain an impediment to this. One of the main problems in modelling species at the global scale is the failure to predict presence of species in enclosed bays/seas where environmental conditions are distinct from surrounding areas (e.g. Red Sea, Baltic Sea and Arafura Sea). If species lack records from such areas then their distributions often reflect this with a predicted absence from the area. Expert review can read-

ily identify these cases and includes various methods for altering species distributions to include these areas.

It will remain the case that an expert in modelling methods should be able to produce a better species distribution model by: using more data (presence and absence or even abundance); using more exhaustive methods (such as Boosted Regression Trees); applying better specific model settings (e.g. regularization in Max-ent); testing for and applying corrections to environmental bias in sampling, and; using more or different environmental layers (potentially at different resolutions) depending on the scale of the analysis. However, such analyses must be made on a case-by-case basis and require good knowledge of both the modelling methods and the species biology, whereas if the aim is to summarise biodiversity generally and quickly (Balmford et al., 2005), species range maps must be produced in the greatest possible number and using a generalised method to maintain consistency. With the datasets available at this time, AquaMaps provides such maps online for a large number of species and to a quality comparable to, if not better than, other methods tested here. In addition it provides the flexibility to review, adapt and store ranges online in a matter of minutes which, as all predictions are in a single system, can then be summarised based on any number of criteria.

Acknowledgements

This work was supported in part by the European Commission under the INCOFISH project (<http://www.incofish.org>), EC INCO contract no. 003739 and in part by Pew Grant UM2004-001034/PPF(NEA)03-01. P. Eastwood and A. South received additional support from the UK Department for Environment, Food, and Rural Affairs through contract AE0916. We thank C. Close at the University of British Columbia for assistance with the preparation of environmental datasets, C. Allison and the Secretariat of the International Whaling Commission for providing us with the IWC catch and sighting datasets, and P. Hammond, Sea Mammal Research Unit, for the SCANS dataset.

References

- Anderson, R.P., Lew, D., Peterson, A.T., 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecol. Model.* 162, 211–232. doi:10.1016/S0304-3800(02)00349-6.
- Balmford, A., Bennun, L., ten Brink, B., Cooper, D., Côté, I.M., Crane, P., Dobson, A., Dudley, N., Dutton, I., Green, R.E., Gregory, R.D., Harrison, J., Kennedy, E.T., Kremen, C., Leader-Williams, N., Lovejoy, T.E., Mace, G., May, R., Mayaux, P., Morling, P., Phillips, J., Redford, K., Ricketts, T.H., Rodriguez, J.P., Sanjayan, M., Schei, P.J., van Jaarsveld, A.S., Walther, B.A., 2005. The convention on biological diversity's 2010 target. *Science* 307, 212–213. doi:10.1126/science.1106281.
- Carr, M.-E., Friedrichs, A.M., Schmeltz, M., Aita, M.N., Antoine, D., Arrigo, K.R., Asanuma, I., Aumont, O., Barber, R., Behrenfeld, M., Bidigare, R., Buitenhuis, E.T., Campbell, J., Ciotti, A., Dierssen, H., Dowell, M., Dunne, J., Esaias, W., Gentili, B., Gregg, W., Groom, S., Hoepffner, N., Ishizaka, J., Kameda, T., Le Quére, C., Lohrenz, S., Marra, J., Mélin, F., Moore, K., Morel, A., Reddy, T.E., Ryan, J., Scardi, M., Smyth, T., Turpie, K., Tilstone, G., Waters, K., Yamanaka, Y., 2006. A comparison of global estimates of marine primary production from ocean color. *Deep-Sea Res.*, II 53, 741–770.
- Cavalieri, D., Parkinson, C., Gloersen, P., Zwally, H.J., 1996 (updated 2006). Sea ice concentrations from Nimbus-7 SMMR and DMSP SSM/I passive microwave data, [1979–2002]. Boulder, Colorado USA: National Snow and Ice Data Center. Digital media.
- CEFAS, 2007. Centre for Environment, Fisheries & Aquaculture Science (Cefas) Web service. Data, Fisheries Information, Surveys <http://www.cefas.co.uk/data/fisheries-information/surveys.aspx>. Accessed December 2007.
- Compagno, L.J.V., 1984. FAO species catalogue Vol.4. Sharks of the world. An Annotated and Illustrated Catalogue of Shark Species Known to Date Part 1 – Hexanchiformes to Lamniformes. FAO Fish. Synop., 125, Part 1.
- Conkright, M.E., Locarnini, R.A., Garcia, H.E., O'Brien, T.D., Boyer, T.P., Stephens, C., Antonov, J.I., 2002. World Ocean Atlas 2001: objective analyses, data statistics, and figures CD-ROM documentation. U.S. National Oceanographic and Atmospheric Administration, National Oceanographic Data Center, Silver Spring, MD, p. 17.
- Ekman, S., 1967. Zoogeography of the Sea. Sidgwick & Jackson, London.

- Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M.C., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129–151.
- Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Env. Cons.* 24 (1), 38–49.
- Froese, R., Pauly, D., 2007. FishBase. World Wide Web electronic publication. <http://www.fishbase.org>. Version 01/2007.
- GBIF, 2007. Global Biodiversity Information Facility. World Wide Web electronic publication <http://www.gbif.org/>. Accessed December 2007.
- Gowans, S., 2002. Bottlenose Whales - *Hyperoodon ampullus* and *H. planifrons*. In: Perrin, W.F., Würsig, B., Theewissen, J.G.M. (Eds.), *Encyclopedia of Marine Mammals*. Academic Press, San Diego, CA, pp. 128–129.
- Grant, W.S., Clark, A.M., Bowen, B.W., 1998. Why restriction fragment length polymorphism analysis of mitochondrial DNA failed to resolve sardine (*Sardinops*) biogeography: insights from mitochondrial DNA cytochrome b sequences. *Can. J. Fish Aquat. Sci.* 55 (12), 2539–2547.
- Guisan, A., Thuiller, W., 2005. Predicting species distribution: offering more than simple habitat models. *Ecol. Lett.* 8, 993–1009.
- Hammond, P.S., Berggren, P., Benke, H., Borchers, D.L., Collet, A., Heide-Jorgensen, M.P., Heimlich, S., Hiby, A.R., Leopold, M.F., Øien, N., 2002. Abundance of harbour porpoise and other cetaceans in the North Sea and adjacent waters. *J. Appl. Ecol.* 39, 361–376.
- Hastie, T.J., 1991. Generalized additive models. In: Chambers, J.M., Hastie, T.J. (Eds.), *Statistical Models*. S. Wadsworth & Brooks/Cole, Chapter 7.
- Hernandez, P.A., Graham, C.H., Máster, L.L., Albert, D.L., 2006. The effect of sample size and species characteristics on performance of different species distribution modelling methods. *Ecography* 29, 773–785.
- IWC (International Whaling Commission), 2001. IDCR-DESS SOWER survey data set (1978–2001). IWC, Cambridge.
- Johnson, C.J., Gillingham, M.P., 2004. Mapping uncertainty: sensitivity of wildlife habitat ratings to expert opinion. *J. Appl. Ecol.* 41, 1032–1041.
- Kasamatsu, F., Matsuoka, K., Hakamada, T., 2000. Interspecific relationships in density among the whale community in the Antarctic. *Polar Biol.* 23, 466–473.
- 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. *MEPS* 316, 285–310.
- Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecol. Biogeogr.* 17, 145–151.
- MacLeod, C.D., D'Amico, A., 2006. A review of beaked whale behaviour and ecology in relation to assessing and mitigating impacts of anthropogenic noise. *J. Cetacean Res. Manage.* 7, 211–221.
- MacLeod, C.D., Mandelberg, L., Schweder, C., Bannon, S.M., Pierce, G.J., 2008. A comparison of approaches for modelling the occurrence of marine animals. *Hydrobiologia* 612, 21–32, doi:10.1007/s10750-008-9491-0.
- MarLIN, 2007. CSIRO Marine and Atmospheric Research (CMAR) Laboratories Information Network <http://www.marine.csiro.au/marlin/> Accessed December 2007.
- McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models*. Chapman and Hall.
- McPherson, J.M., Jetz, W., Rogers, D.J., 2004. The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *J. Appl. Ecol.* 41, 811–823.
- Meier, R., Dikow, T., 2004. Significance of specimen databases from taxonomic revisions for estimating and mapping the global species diversity of invertebrates and repatriating reliable specimen data. *Cons. Biol.* 18 (2), 478–488.
- Muñoz, M.E.S., Scachetti-Pereira, R., De Giovanni, R., Sutton, T., Ruland, K., Brewer, P., Oberender, J., Elwertowski, T., Jardim, A.C., Bellini, D.J.S., Cheng-Tao, L., 2007. openModeller. <http://openmodeller.sf.net>. Version 1.0.4 Accessed 26/01/07.
- NOAA, 2006. ETOPO2 v2 global gridded 2-minute database. U.S. National Oceanic and Atmospheric Administration, National Environmental Satellite, Data, and Information Service, National Geophysical Data Center. <http://www.ngdc.noaa.gov/mgg/image/2minrelief.html>. Accessed December 2007.
- NOAA, 2007. NOAA optimum interpolation sea surface temperature analysis. U.S. National Oceanic and Atmospheric Administration, National Centre for Environmental Prediction. <http://www.emc.ncep.noaa.gov/research/cmb/sst.analysis>. Accessed December 2007.
- OBIS, 2007. Ocean Biogeographic Information System. <http://www.iobis.org>. Accessed December 2007.
- Pearce, J.L., Boyce, M.S., 2006. Modelling distribution and abundance with presence-only data. *J. Appl. Ecol.* 43, 405–412.
- Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modelling of species geographic distributions. *Ecol. Model.* 190, 231–259.
- Phillips, S.J., 2008. Transferability, sample selection bias and background data in presence-only modelling: a response to Peterson et al. (2007). *Ecography Online* Early doi:10.1111/j.2007.0906-7590.05378.x.
- Rees, T., 2003. C-squares, a new spatial indexing system and its applicability to the description of oceanographic datasets. *Oceanography* 16, 11–19.
- Reynolds, R.W., Smith, T.M., 1995. A high-resolution global sea surface temperature climatology. *J. Climate* 8, 1571–1583.
- Rogers, S.I., 1992. Environmental factors affecting the distribution of sole (*Solea solea* (L.)) within a nursery area. *Neth. J. Sea Res.* 29, 153–161.
- Sing, T., Sander, O., Beerwinkel, N., Lengauer, T., 2005. ROCr: visualizing classifier performance in R. *Bioinformatics* 21, 3940–3941.
- Siqueira, M.F., Durigan, G., 2007. Modelagem da distribuição geográfica de espécies lenhosas de cerrado no Estado de São Paulo. *Revista Brasileira de Botânica* 30, 239–249.
- Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. *Ecol. Model.* 148, 1–13.
- Swets, J.A., 1988. Measuring the accuracy of diagnostic systems. *Science* 240, 1285–1293.