Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model

K. Kaschner1,2,3,*, R. Watson1, A. W. Trites2, D. Pauly1

1Sea Around Us Project, Fisheries Centre, University of British Columbia, 2259 Lower Mall, Vancouver, British Columbia V6T 1Z4, Canada
2Marine Mammal Research Unit, Fisheries Centre, University of British Columbia, Hut B-3, 6248 Biological Sciences Road, Vancouver, British Columbia V6T 1Z4, Canada
3Forschungs- und Technologiezentrum Westküste, Hafentörn, 25761 Büsum, Germany

ABSTRACT: The lack of comprehensive sighting data sets precludes the application of standard habitat suitability modeling approaches to predict distributions of the majority of marine mammal species on very large scales. As an alternative, we developed an ecological niche model to map global distributions of 115 cetacean and pinniped species living in the marine environment using more readily available expert knowledge about habitat usage. We started by assigning each species to broad-scale niche categories with respect to depth, sea-surface temperature, and ice edge association based on synopses of published information. Within a global information system framework and a global grid of 0.5° latitude/longitude cell dimensions, we then generated an index of the relative environmental suitability (RES) of each cell for a given species by relating known habitat usage to local environmental conditions. RES predictions closely matched published maximum ranges for most species, thus representing useful, more objective alternatives to existing sketched distributional outlines. In addition, raster-based predictions provided detailed information about heterogeneous patterns of potentially suitable habitat for species throughout their range. We tested RES model outputs for 11 species (northern fur seal, harbor porpoise, sperm whale, killer whale, hourglass dolphin, fin whale, humpback whale, blue whale, Antarctic minke, and dwarf minke whales) from a broad taxonomic and geographic range, using data from dedicated surveys. Observed encounter rates and species-specific predicted environmental suitability were significantly and positively correlated for all but 1 species. In comparison, encounter rates were correlated with <1% of 1000 simulated random data sets for all but 2 species. Mapping of large-scale marine mammal distributions using this environmental envelope model is helpful for evaluating current assumptions and knowledge about species occurrences, especially for data-poor species. Moreover, RES modeling can help to focus research efforts on smaller geographic scales and usefully supplement other, statistical, habitat suitability models.

KEY WORDS: Habitat suitability modeling · Marine mammals · Global · GIS · Relative environmental suitability · Niche model · Distribution

INTRODUCTION

A number of marine mammal species are currently threatened by a variety of anthropogenic factors, ranging from bycatch and ship-strikes to pollution, global warming, and potential food competition (Perrin et al. 2002). The development and implementation of effective conservation measures require, however, detailed knowledge about the geographic occurrence of a species. In recent years, advances in geographic information systems (GIS) and computational power have allowed the development and application of habitat
suitability models to quantitatively delineate maximum range extents and predict species’ distributions. Standard models rely on available occurrence records to investigate the relationships between observed species’ presence and the underlying environmental parameters that—either directly or indirectly—determine a species’ distribution in a known area and use this information to predict the probability of a species’ occurrence in other areas (Guisan & Zimmermann 2000).

Habitat suitability models have been widely applied in terrestrial systems and for a wide range of land-based species (Peterson & Navarro-Sigüenza 1999, Zaniewski et al. 2002, Store & Jokimäki 2003). There are, however, comparatively few attempts to use such models to map species’ distributions in the marine environment (Huettmann & Diamond 2001, Yen et al. 2004, Guinotte et al. 2006 in this Theme Section). This is particularly true for marine mammals, partly because the collection of species’ occurrence data is hampered by the elusiveness and mobility of these animals. In addition, designated and costly surveys usually cover only a small fraction of a species’ range (e.g. Kasamatsu et al. 2000, Hammond et al. 2002, Waring et al. 2002), due to the vastness of the marine environment and the panglobal distributions of many species. Thus, these surveys often yield little more than a snapshot, both in time and space, of a given species’ occurrence. The comparatively low densities of many marine mammal species further contribute to the difficulties in distinguishing between insufficient effort to detect a species in a given area and its actual absence. On the other hand, a concentration of sightings may only reflect the concentration of effort rather than a concentration of occurrence (Kenney & Winn 1986).

There are on-going efforts—conducted, for example, as part of the OBIS initiative (Ocean Biogeographic Information System)—to compile existing marine mammal occurrence records, to allow for large-scale quantitative analyses of species distributions using habitat suitability modeling. For many species, however, there have been <12 known or published sightings to date. Actual point data sets, which generally cover only a fraction of known range extents, are available or readily accessible for <50% of all marine mammal species through the OBIS-SEAMAP portal (http://seamap.env.duke.edu/), the currently most comprehensive data repository for marine mammal sightings.

As a consequence of this data paucity, marine mammal occurrence has been modeled for only a handful of species and only in relatively small areas. Most existing studies have employed so-called presence–absence statistical models, such as general linear models (GLMs) or general additive models (GAMs) (Moses & Finn 1997, Hedley et al. 1999, Gregr & Trites 2001, Hamazaki 2002). These model types require data collected during line-transect surveys that systematically document species’ presences and absences to predict varying species’ densities or probabilities of occurrence (Hamazaki 2002, Hedley & Buckland 2004). However, predictions from presence–absence type models are affected by species’ prevalence (Manel et al. 2001). For marine mammals, however, densities and/or detectability tend to be very low. More importantly, representative survey coverage of entire range extents has currently been achieved for an estimated 2% of all species. This precludes the application of presence–absence modeling techniques to predict occurrence on larger scales for the vast majority of all cetaceans and pinnipeds.

Ecological niche models such as GARP (Genetic Algorithm for Rule Set Production; Stockwell & Noble 1992) and ecological niche factor analysis (ENFA) (Hirzel et al. 2002) represent alternative approaches which—due to their more mechanistic nature—can reduce the amount of data needed, since they do not require absence data and may therefore use so-called opportunistic data sets. These presence-only models have found widespread application in terrestrial systems (Peterson et al. 2000, Peterson 2001, Engler et al. 2004), and, more recently, attempts have been made to use such models to predict distributions of some rarer marine mammal species (Compton 2004, MacLeod 2005). However, for most species, there are fewer occurrence records readily available than required to generate accurate predictions (e.g. 50 to 100 representative occurrence records in the case of GARP, Stockwell & Peterson 2002). Moreover, these niche models assume that data sets represent an unbiased sample of the available habitat (Hirzel et al. 2002), which makes them sensitive to the skewed distribution of effort prevalent in most opportunistically collected marine mammal data sets (see below).

In conclusion, the current shortage of point data sets has prevented applying standard empirical habitat suitability models to predict patterns of occurrences or maximum range extents on larger scales. Similarly, this lack of data has prohibited the prediction of occurrence patterns for the lesser-known marine mammal species in more inaccessible or understudied regions of the world’s oceans—and will likely continue to do so in the foreseeable future. As a consequence, marine mammal distributional ranges published to date mainly consist of hand-drawn maps outlining the proposed maximum area of a species’ occurrence based on the professional judgment of experts and synopses of qualitative information (e.g. Ridgway & Harrison 1981a,b, 1985, 1989, 1994, 1999, Perrin et al. 2002). Frequently, there is considerable variation amongst the range extents proposed by different authors for the same species (Jefferson et al. 1993, Reijnders et al. 1993). In addition, these maps are often supplemented
by relatively large regions covered by question marks, indicating areas of unknown, but likely, occurrence. As an alternative, some authors have summarized available raw point data in the form of documented stranding or sighting locations on maps (e.g. Perrin et al. 1994, Jefferson & Schiro 1997, Ballance & Pitman 1998), thus leaving it to the readers to infer possible species’ distributions. All of these approaches are greatly confounded by uncertainty in the degree of interpolation applied to the occurrence data (Gaston 1994), and none delineates species’ distributions based on an explicit algorithm that captures patterns of species’ occurrences using a rule-based approach or statistical models, as recommended by Gaston (1994).

Although we currently lack the comprehensive point data sets to remedy this situation using standard habitat suitability modeling techniques, we nevertheless already know quite a bit about the general habitat usage of most marine mammal species, available in the form of qualitative descriptions, mapped outlines, graphically fragmented quantitative observations, and large-scale historical catch data sets. Existing knowledge about species’ occurrence is likely biased — given the high concentration of survey efforts in shelf waters of the northern hemisphere — and the lack of statistical investigations on resource selection does not allow definitive conclusions about habitat preferences for most species (Johnson 1980, Manly et al. 2002). However, the synthesis of available knowledge about species’ occurrences, collected from wide range of sources, time periods, and geographic regions, may approximate a representative sampling scheme in terms of the investigation of habitat usage on very large scales — at least until sufficient point data sets become available for more rigorous analyses. In the meantime, we propose that expert knowledge may represent an alternative and underutilized resource that can form the basis for the development of other types of habitat suitability models, such as rule-based environmental envelope models. Envelope models and techniques relying on formalized expert opinion have frequently been used in the past to predict large-scale terrestrial plant distributions (e.g. Shao & Halpin 1995, Guisan & Zimmerman 2000, Skov & Svenning 2004), but have not yet been applied to describe marine mammal range extents.

The objective of this study was to develop a generic quantitative approach to predict the average annual geographical ranges of all marine mammal species within a single conceptual framework using basic descriptive data that were available for (almost) all species. We also wanted to gain insight into the potential relative environmental suitability (RES) of a given area for a species throughout this range. Since comprehensive point data sets are currently non-existent or non-accessible for the vast majority of marine mammal species, we sought to generate our predictions based on the synthesis of existing and often general qualitative observations about the spatial and temporal relationships between basic environmental conditions and a given species’ presence. The maps we produced represent a visualization of existing knowledge about a species’ habitat usage, processed in a standardized manner within a GIS framework and related to local environmental conditions. Thus, our results can be viewed as hypotheses about potentially suitable habitat or main aspects of a species’ fundamental ecological niche, as defined by Hutchinson (1957). We tested and evaluated our model predictions and assumptions using available marine mammal sightings and catch data from different regions and time periods to establish the extent to which this approach may be able to capture actual patterns of species’ occurrence. Finally, we explored the merits and limitations of the model as a useful supplement to existing habitat suitability modeling approaches.

MATERIALS AND METHODS

Model structure, definitions, scope, and resolution.
We derived the geographic ranges for 115 marine mammal species and predicted the RES for each of them throughout this range based on the available information about species-specific habitat usage. We defined geographic range as the maximum area between the known outer-most limits of a species’ regular or periodic occurrence. While this definition is inclusive of all areas covered during annual migrations, dispersal of juveniles etc., it specifically excludes extralimital sightings, which are sometimes difficult to distinguish from the core range (Gaston 1994). Adhering to the plea of Hall et al. (1997) for the use of clear definitions and standard terminology, we chose the term ‘relative environmental suitability’ rather than ‘habitat suitability’ to describe model outputs, to distinguish our predictions, which often corresponded more closely to a species’ fundamental niche, from the actual probabilities of occurrence generated by other habitat suitability models (Hirzel et al. 2002).

General patterns of occurrence of larger, long-living animals, such as marine mammals, are unlikely to be affected by environmental heterogeneity over small temporal and spatial scales (Turner et al. 1995, Jaquet 1996). This may be especially true for species living in the marine environment, as pelagic systems show greater continuity in environmental conditions over evolutionary time than terrestrial environments (Platt & Sathyendranath 1992). We chose a global geographic scope to accommodate the wide-ranging
annual movements and cosmopolitan occurrence of numerous marine mammal species. Similarly, we used long-term averages of temporally varying environmental parameters to minimize the impacts of interannual variation. The model’s spatial grid resolution of 0.5° latitude by 0.5° longitude represents a widespread standard for global models.

**Independent variables.** The lack of point data used for model input precluded the application of standard techniques to determine which environmental predictors might be best suited to predict species’ occurrence. Instead, selection of environmental proxies that served as independent variables in our model was based on the existing knowledge about their relative importance to—indirectly—determine species occurrence for many marine mammals. Furthermore, predictors were chosen based on the availability of data at appropriate scales, including the availability of matching habitat usage information that was obtainable for all or at least the majority of all species. All environmental data were interpolated and rasterized using a custom GIS software package (SimMap 3.1 developed by R. Watson & N. Hall) and stored as attributes of individual grid cells in the global raster (Watson et al. 2004) (Fig. 1A–C).

**Bottom depth:** Strong correlations between bathymetry and patterns of inter- or intraspecific species’ occurrences have been noted for many species of cetaceans and pinnipeds in different regions and ocean basins (Payne & Heinemann 1993, Moore et al. 2002, Baumgartner et al. 2001, Hamazaki 2002), making seafloor elevation an ideal candidate as an environmental proxy for a generic habitat suitability model. Bathymetric data were taken from the ETOPO2 dataset available on the United States National Geophysical Data Center’s ‘Global Relief’ CD (www.ngdc.noaa.gov/products/ngdc_products.html), which provides elevation in 2 min intervals for all points on earth (Fig. 1A).

**Mean annual sea-surface temperature:** In addition to non-dynamic parameters, such as bathymetry, marine mammal distributions are influenced by a host of variable environmental factors, such as sea-surface temperature (SST). Changes in SST may be indicative of oceanographic processes that ultimately determine marine mammal occurrence across a number of different temporal scales (Au & Perryman 1985), and significant correlations of marine mammal species with SST have been demonstrated in different areas and for a variety of different species (e.g. Davis et al. 1998, Baumgartner et al. 2001, Hamazaki 2002). Surface temperature may not be a good predictor for all marine mammals, given the substantial foraging depths of some species (Jaquet 1996). However, we nevertheless chose to use SST as a proxy, because of the general availability of observations of surface climatic conditions or quantitative measurements associated with marine mammal occurrences. Global annual SST data, averaged over the past 50 yr, were extracted from the NOAA World Ocean Atlas 1998 CD (NOAA/NODC 1998) (Fig. 1B).

**Mean annual distance to ice edge:** The shifting edge of the pack ice is a highly productive zone (Brierley et al. 2002, Hewitt & Lipsky 2002) and represents important feeding grounds for many species of marine mammals (Murase et al. 2002). A number of studies have shown that sea ice concentration and ice cover, in combination with depth, play a key role in ecological niche partitioning for many species (Ribic et al. 1991, Moore & DeMaster 1997). We included the distance to the ice edge as an additional predictor in our model, as the distribution of species in the polar zones may not be fully captured using only SST. Although ice extent is strongly spatially correlated with SST, the actual edge of the sea ice does not directly coincide with any single isotherm throughout the year (Fig. 1B,C). Moreover, the ability of different marine mammal species to venture into pack-ice varies substantially. Spatial information about the average monthly ice extent (1979 to 1999)—defined by the border of minimum 50% sea ice coverage—was obtained from the United States National Snow & Ice Data Center (NSIDC) website (http://nsidc.org/data/smr_ssmi_ancillary/trends.html#gis). We smoothed the ice edge border to correct some obvious misclassification and/or re-projection errors. After rasterizing the ice extent data, we calculated monthly distances from the nearest ice edge cell for each cell in the raster and computed annual average distances based on these monthly distances (Fig. 1C).

**Distance to land:** Some pinniped species—specifically the eared seals (otariids)—appear to be restricted to areas fairly close to their terrestrial resting sites, i.e. haulouts and rookeries (Costa 1991, Boyd 1998). The maximum distances away from these land sites are determined by a combination of species-specific life-history and physiological factors, such as the maximum nursing intervals based on the ability of pups to fast (Bonner 1984) and maximum swimming speed of adults (Ponganis et al. 1992). Global data sets identifying pinniped rookery sites do not exist. However, distance from landmasses in general was deemed to be an appropriate proxy in the context of this model and
served as an additional predictor to more realistically model the distribution of some of the pinniped species (Appendix 2 in Kaschner 2004). For each cell, distance to land, defined as the nearest cell containing a part of coastline, was calculated in the same manner as distance to the ice edge.

**Dependent variables.** Marine mammal species: Our model encompassed 115 species of marine mammals that live predominantly in the marine environment (Table 1, present paper, and Appendix 1 in Kaschner 2004). We did not consider exclusively freshwater cetaceans or pinnipeds, nor the marine sireniens, sea otters, or the polar bear. Taxonomically, we largely followed Rice (1998), except for right whales, for which we recognized 3 separate species (Rosenbaum et al. 2000, Bannister et al. 2001). In addition, we included a recently described additional species, Perrin's beaked whale *Mesoplodon perrini* (Dalebout et al. 2002).

**Definition of habitat usage or niche categories:** Habitat usage categories were defined to represent broad predictor ranges, which roughly describe real marine physical/ecological niches inhabited by different marine mammal species. Niche categories effectively represent species response curves in relation to available habitat. Normally such response curves are derived empirically based on the statistical analysis of animal occurrences in relation to direct or indirect ecological gradients (Guisan & Zimmermann 2000, Manly et al. 2002). However, again, for the vast majority of marine mammal species the possible shape of such relationships remains to be investigated, and in the few existing studies only a sub-set of the available habitat has been covered (e.g. Cañadas et al. 2003).

The more mechanistic nature of our model and the non-point type input data used precluded the derivation of empirical generic relationships within the context of this study. We therefore assumed a trapezoidal response curve (Fig. 2). We selected this shape as the most broadly appropriate option to model annual average distributions, as it represents a compromise between the likely unimodal response curves for species with fairly restricted ranges and the probably more bi-modal shape for species undertaking substantial migrations. The selected shape meant that the relative environmental suitability was assumed to be uniformly highest throughout a species’ preferred or mostly used parameter range (MinP to MaxP in Fig. 2). Beyond this range, we assumed that suitability would generally decrease linearly towards the minimum or maximum thresholds for a species (MinA or MaxA in Fig. 2). Suitability was set to zero outside the absolute minimum or maximum values.

While ecologically meaningful niches for bottom depth and association with ice extent are variable in

**Table 1.** Names, taxonomy, and general distributions of the 20 selected marine mammal species included in the relative environmental suitability (RES) model for which we show predictions (see Fig. 3) (for all other species see Kaschner 2004, her Appendix 1)

<table>
<thead>
<tr>
<th>Common name</th>
<th>Scientific name</th>
<th>Suborder</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Atlantic right whale</td>
<td><em>Balaena glacialis</em></td>
<td>Mysticeti</td>
<td>N Atlantic</td>
</tr>
<tr>
<td>Antarctic minke whale</td>
<td><em>Balaenoptera bonaerensis</em></td>
<td>Mysticeti</td>
<td>S hemisphere</td>
</tr>
<tr>
<td>Gray whale</td>
<td><em>Eschrichtius robustus</em></td>
<td>Mysticeti</td>
<td>N Pacific</td>
</tr>
<tr>
<td>Hourglass dolphin</td>
<td><em>Lagenorhynchus cruciger</em></td>
<td>Odontoceti</td>
<td>S hemisphere</td>
</tr>
<tr>
<td>Northern right whale dolphin</td>
<td><em>Lissodelphis borealis</em></td>
<td>Odontoceti</td>
<td>N Pacific</td>
</tr>
<tr>
<td>Irrawaddy dolphin</td>
<td><em>Orcaella brevirostris</em></td>
<td>Odontoceti</td>
<td>Indo-Pacific</td>
</tr>
<tr>
<td>Indian hump-backed dolphin</td>
<td><em>Sousa plumbea</em></td>
<td>Odontoceti</td>
<td>W Indian Ocean</td>
</tr>
<tr>
<td>Clymene dolphin</td>
<td><em>Stenella clymene</em></td>
<td>Odontoceti</td>
<td>Atlantic</td>
</tr>
<tr>
<td>Narwhal</td>
<td><em>Monodon monoceros</em></td>
<td>Odontoceti</td>
<td>Circumpolar, N hemisphere</td>
</tr>
<tr>
<td>S African &amp; Australian fur seal</td>
<td><em>Arctocephalus pusillus</em></td>
<td>Pinnipedia</td>
<td>S Africa, S Australia</td>
</tr>
<tr>
<td>Guadalupe fur seal</td>
<td>A. townsendi</td>
<td>Pinnipedia</td>
<td>NE Pacific</td>
</tr>
<tr>
<td>New Zealand fur seal</td>
<td>A. forsteri</td>
<td>Pinnipedia</td>
<td>New Zealand, S Australia</td>
</tr>
<tr>
<td>Australian sea lion</td>
<td><em>Neophoca cinerea</em></td>
<td>Pinnipedia</td>
<td>S America</td>
</tr>
<tr>
<td>South (American) sea lion</td>
<td><em>Otaria flavescens</em></td>
<td>Pinnipedia</td>
<td>Galapagos Islands, E Pacific</td>
</tr>
<tr>
<td>Galapagos sea lion</td>
<td><em>Zalophus wollebaeki</em></td>
<td>Pinnipedia</td>
<td>N Atlantic</td>
</tr>
<tr>
<td>Hooded seal</td>
<td><em>Cystophora cristata</em></td>
<td>Pinnipedia</td>
<td>N Pacific</td>
</tr>
<tr>
<td>Ribbon seal</td>
<td><em>Histriophoca fasciata</em></td>
<td>Pinnipedia</td>
<td>Mediterranean, NE Atlantic</td>
</tr>
<tr>
<td>Mediterranean monk seal</td>
<td><em>Monachus monachus</em></td>
<td>Pinnipedia</td>
<td>Hawaii, NE Pacific</td>
</tr>
<tr>
<td>Hawaiian monk seal</td>
<td><em>M. schauinslandi</em></td>
<td>Pinnipedia</td>
<td>Circumpolar, S hemisphere</td>
</tr>
<tr>
<td>Ross seal</td>
<td><em>Ommatophoca rossii</em></td>
<td>Pinnipedia</td>
<td></td>
</tr>
</tbody>
</table>
width and were defined accordingly, SST categories were described by regular 5°C steps, based on the average intra-annual variation of 5 to 10°C in most areas of the world (Angel 1992). Quantitative definitions and corresponding qualitative descriptions of potential niches of the resulting 17 bottom depth ranges, 28 broad temperature ranges, and 12 ice edge association categories are shown in Table 2.

Table 2. Quantitative and qualitative definitions of habitat usage or niche categories (SST: sea-surface temperature; cont.: continental)

<table>
<thead>
<tr>
<th>Environmental parameter</th>
<th>Minimum</th>
<th>Preferred</th>
<th>Maximum</th>
<th>Habitat category description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth usage zones (in m)</td>
<td></td>
<td></td>
<td></td>
<td>All depths (uniform distribution)</td>
</tr>
<tr>
<td></td>
<td>0 –1</td>
<td>0 –8000</td>
<td>0 –8000</td>
<td>Mostly estuarine to edge of cont. shelf</td>
</tr>
<tr>
<td></td>
<td>0 –50</td>
<td>0 –200</td>
<td>0 –200</td>
<td>Mainly coastal–upper cont. shelf to upper cont. slope</td>
</tr>
<tr>
<td></td>
<td>0 –10</td>
<td>0 –100</td>
<td>0 –100</td>
<td>Mainly coastal–upper cont. shelf to upper cont. slope</td>
</tr>
<tr>
<td></td>
<td>0 –200</td>
<td>0 –6000</td>
<td>0 –6000</td>
<td>Mainly coastal–cont. shelf to deep waters</td>
</tr>
<tr>
<td></td>
<td>0 –1000</td>
<td>0 –6000</td>
<td>0 –6000</td>
<td>Mainly coastal–upper cont. slope to deep waters</td>
</tr>
<tr>
<td></td>
<td>0 –2000</td>
<td>0 –8000</td>
<td>0 –8000</td>
<td>Mainly coastal–cont. slope to very deep waters</td>
</tr>
<tr>
<td></td>
<td>0 –4000</td>
<td>0 –8000</td>
<td>0 –8000</td>
<td>Mainly coastal–abyssal plains to very deep waters</td>
</tr>
<tr>
<td></td>
<td>0 –200</td>
<td>0 –6000</td>
<td>0 –6000</td>
<td>Mainly upper cont. shelf to deep waters</td>
</tr>
<tr>
<td></td>
<td>0 –2000</td>
<td>0 –6000</td>
<td>0 –6000</td>
<td>Mainly upper cont. slope to deep waters</td>
</tr>
<tr>
<td></td>
<td>0 –2000</td>
<td>0 –8000</td>
<td>0 –8000</td>
<td>Mainly upper cont. slope to very deep waters</td>
</tr>
<tr>
<td></td>
<td>0 –1000</td>
<td>0 –8000</td>
<td>0 –8000</td>
<td>Mainly lower cont. slope to very deep waters</td>
</tr>
<tr>
<td></td>
<td>0 –4000</td>
<td>0 –8000</td>
<td>0 –8000</td>
<td>Mainly lower cont. slope–abyssal plains to very deep waters</td>
</tr>
<tr>
<td></td>
<td>0 –200</td>
<td>0 –6000</td>
<td>0 –6000</td>
<td>Mainly abyssal plains to very deep waters</td>
</tr>
<tr>
<td>Temperature usage zones</td>
<td>−2</td>
<td>−2</td>
<td>35</td>
<td>All temperatures (uniform distribution)</td>
</tr>
<tr>
<td>(mean annual SST, in °C)</td>
<td>−2</td>
<td>0</td>
<td>3</td>
<td>Polar only</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>Polar–subpolar</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>5</td>
<td>10</td>
<td>Polar–cold temperate</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>10</td>
<td>15</td>
<td>Polar–warm temperate</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>15</td>
<td>20</td>
<td>Polar–tropical</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>30</td>
<td>35</td>
<td>Polar–full tropical</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>Subpolar only</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>5</td>
<td>10</td>
<td>Subpolar–cold temperate</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10</td>
<td>15</td>
<td>Subpolar–warm temperate</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>15</td>
<td>20</td>
<td>Subpolar–tropical</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>20</td>
<td>25</td>
<td>Subpolar–subtropical</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>25</td>
<td>30</td>
<td>Subpolar–tropical</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>30</td>
<td>35</td>
<td>Subpolar–tropical</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10</td>
<td>15</td>
<td>Cold temperate only</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10</td>
<td>20</td>
<td>Cold temperate–warm temperate</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10</td>
<td>25</td>
<td>Cold temperate–subtropical</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10</td>
<td>30</td>
<td>Cold temperate–tropical</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>10</td>
<td>35</td>
<td>Cold temperate–full tropical</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>15</td>
<td>15</td>
<td>Warm temperate only</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>15</td>
<td>20</td>
<td>Warm temperate–cold temperate</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>15</td>
<td>30</td>
<td>Warm temperate–tropical</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>15</td>
<td>30</td>
<td>Warm temperate–full tropical</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>20</td>
<td>20</td>
<td>Subtropical only</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>Subtropical only</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>20</td>
<td>30</td>
<td>Subtropical–cold temperate</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>20</td>
<td>30</td>
<td>Subtropical–tropical</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>25</td>
<td>20</td>
<td>Subtropical–tropical</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>Subtropical–tropical</td>
</tr>
<tr>
<td>Ice edge usage zones</td>
<td>−1</td>
<td>0</td>
<td>8000</td>
<td>No association with ice edge (uniform distribution)</td>
</tr>
<tr>
<td>(mean annual distance</td>
<td>−1</td>
<td>0</td>
<td>8000</td>
<td>Mostly restricted to fast &amp; deep pack-ice</td>
</tr>
<tr>
<td>from ice edge, in km)</td>
<td>−1</td>
<td>0</td>
<td>8000</td>
<td>Mainly in fast &amp; deep pack-ice, but also elsewhere</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>500</td>
<td>Mainly around edge of pack-ice</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>500</td>
<td>Mainly around edge of pack-ice, but also elsewhere</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>8000</td>
<td>Mainly in areas of max. ice extent, but also elsewhere</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>8000</td>
<td>Regularly but not preferably around edge of the pack-ice</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>500</td>
<td>8000</td>
<td>Mainly in areas of max. ice extent, but also elsewhere</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>500</td>
<td>8000</td>
<td>Regularly but not preferably in areas of max. ice extent</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1000</td>
<td>8000</td>
<td>Regularly but not preferably close to max. ice extent</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>1000</td>
<td>8000</td>
<td>No association with ice edge, nowhere near ice at any time of the year</td>
</tr>
</tbody>
</table>
Marine mammal habitat usages: We compiled published information about species-specific habitat usages with respect to their known association with the ice edge, as well as commonly inhabited bottom depth and SST ranges. Where appropriate, additional information about maximum likely distance from landmasses was also collected, based on information about maximum foraging trip lengths. Selected sources of information included >1000 primary and secondary references, all screened for relevant information on habitat use (compiled in Kaschner 2004, Appendix 2). Data extracted from these sources ranged from statistically significant results of quantitative investigations of correlations between species’ occurrence and environmental predictors (e.g. Gregr & Trites 2001, Moore et al. 2002, Baumgartner et al. 2003, Cañadas et al. 2003), opportunistic observations (e.g. Carlström et al. 1997), maps of sightings or distribution outlines, to qualitative broad descriptions of prevalent occurrence such as ‘oceanic, subtropical species’ (e.g. Jefferson et al. 1993). A level of confidence was assigned to each record to reflect the origin, reliability, and detail of the data, with quantitative investigations of environmental factors and species’ occurrence ranking highest and qualitative descriptions ranking lowest.

We assigned each species to niche categories for depth, temperature, and ice-edge association (and in some cases distance to land) based on the most reliable information available (Table 3, present paper, and Kaschner 2004, Appendix 2). If the available information was inconclusive, or different conclusions could be drawn from the data, the species was assigned to multiple alternative niche categories representing different hypotheses. Distance from land preferences were used as an additional constraining factor for all species marked by an asterisk in Table 3 (present paper) and in Appendix 2 (Kaschner 2004). For a few species (<5), the general temperature categories were adjusted to reflect the extreme narrowness of their niche.

Area restrictions: On a global scale, contemporary distributions of marine mammals and other species are the result of their evolutionary history. Present occurrences and restrictions to certain areas therefore reflect a species center of origin and ability to disperse defined by its ecological requirements and competitors (LeDuc 2002, Martin & Reeves 2002). Information about a species’ restriction to large ocean basins (i.e. North Atlantic or southern hemisphere), therefore, served as a rough first geographical constraint in the RES prediction model for each species to capture the results of this evolutionary process. The restriction to general ranges corresponds to the first-order selection of species in terms of habitat usage as described by Johnson (1980), and is implicitly incorporated in the sampling designs of many investigations of species’ occurrence (Buckland et al. 1993).

If generated RES predictions did not reflect documented species’ absences from certain areas, further geographical restrictions were imposed (Table 3, ‘excluded areas’). It should be noted, however, that such restrictions were only imposed when known areas of non-occurrence were clearly definable, such as ‘marginal’ ocean basins (e.g. Red, Mediterranean, or Baltic Seas) or RES predictions showed signs of bi- or multimodality, meaning that areas of high suitability were separated by long stretches of less suitable habitat. We minimized introductions of such additional constraints so as not to impede the assessment of the ability of the RES model to describe, on its own, patterns of species’ presence and absence.

Model algorithm—resource selection function. In our global raster, we generated an index of species-specific relative environmental suitability of each individual grid cell by scoring how well its physical attributes matched what is known about a species’ habitat use. RES values ranged between 0 and 1 and represented the product of the suitability scores assigned to the individual attributes (bottom depth, SST, distance from the ice edge, and, in some cases, from land), which were calculated using the assumed trapezoidal response curves described above. A multiplicative approach was chosen to allow each predictor to serve as an effective ‘knock-out’ criterion (i.e. if a cell’s average depth exceeded the absolute maximum of a species’ absolute depth range, the overall RES should be zero, even if annual STT and distance to ice edge of the cell were within the species preferred or overall habitat range).

Multiple hypotheses about species distributions were generated using different combinations of predictor category settings if a species had been assigned to multiple, equally plausible, options of niche categories based on available data. The lack of test data sets for most species precluded the application of standard model evaluation techniques to determine the best model fit (Fielding & Bell 1997). Consequently, we selected the hypothesis considered to represent the best model fit through an iterative process and by qualitative comparison of outputs with all available information about the species’ distribution and occurrence patterns within its range. Objective geographic ranges of species can then be determined based on some pre-defined threshold of predicted low or non-suitability of areas for a given species.

Model evaluation—species response curves and impact of effort biases. To assess the validity of using the RES model instead of available presence-only models, we investigated the degree to which available opportunistic data sets—for species with global or semi-
global distributions—may meet the basic assumption of existing niche models, i.e. unbiased effort coverage. The commercial whaling data is one of the largest opportunistic data sets of marine mammal occurrence, spanning almost 200 yr and approximating global coverage. Whaling operations did not adhere to any particular sampling schemes, and effort distributions were likely strongly biased. Nevertheless, it has been argued that such long-term catch data sets may still serve as good indicators of annual average species distribution and may thus provide some quantitative insight into general patterns of occurrence (Whitehead & Jaquet 1996, Gregr 2000). Consequently, whaling data would seem to be an obvious candidate for predicting distributions of marine mammal species with cosmopolitan or quasi-cosmopolitan range extents using existing presence-only modeling techniques. Using this data, we wanted to assess potential effort biases by comparing large-scale species response curves to environmental gradients derived from opportunistic and non-opportunistic data sets. In addition, we wanted to use the obtained response curves to evaluate the generic trapezoidal shape of our niche categories and how well habitat usage deduced from point data would correspond to the general current knowledge about such usages of specific species, as represented by the assigned niche category.

The opportunistically collected whaling data set contained commercial catches of member states of the International Whaling Commission (IWC) between 1800 and 2001 and was compiled by the Bureau of International Whaling Statistics (BIWS) and the Museum of Natural History, London, UK (IWC 2001a). We analyzed whaling data following an approach similar to that taken by Kasamatsu et al. (2000) and Cañadas et al. (2002) when investigating cetacean occurrence in relation to environmental gradients and generated species’ response curves for 5 species with quasi-cosmopolitan distributions, including sperm whales Physeter macrocephalus, blue whales Balaenoptera musculus, fin whales Balaenoptera physalus, humpback whales Megaptera novaeangliae, and dwarf minke whales B. acutorostrata. The dwarf minke whale occurs to some extent sympatrically with its closely related sister species, the Antarctic minke whale B. bonaerensis. However, the 2 species are generally not distinguished in most data sets, and the analysis conducted therefore relates to a generic minke whale. As a first step, we assigned all catches recorded with accurate positions to the corresponding cell in our global raster, thus obtaining information about mean depth, SST, and distance to ice edge associated with each catch position. We then plotted frequency distributions of globally available habitat and the amount of habitat covered by whaling effort as the percent of total cells falling into each environmental stratum (defined to correspond to breakpoints in our niche categories) for depth, SST, and ice edge distance, to assess the extent to which whalers may have sampled a representative portion of the habitat available to species with global distributions.

To further assess potential effort biases, we generated histograms of catch ‘presence’ cells for individual species. These were based on the number of cells for which any catch of a specific species was reported within an environmental stratum and essentially represent visualizations of this species’ response curve in relation to an environmental gradient. We then compared histograms based on catch ‘presence’ cells with both encounter rate distributions obtained from a non-opportunistic data set and catch distributions corrected for effort using an effort proxy developed during this study.

The non-opportunistic data set was collected during the IDCR/SOWER line-transect surveys, conducted annually over the past 25 yr in Antarctic waters and stored in the IWC-DESS database (IWC 2001b). Similar to the treatment of whaling data, we binned sighting records by raster cells, using only those records with sufficient spatial and taxonomic accuracy (i.e. sighting positions of reliably identified species were reported to, at least, the nearest half degree latitude or longitude). We then calculated species-specific encounter rates or SPUEs (sightings per unit of effort) across all years by computing total length of on-effort transects within each cell using available information about transect starting and end points. Finally, we plotted average SPUEs per environmental stratum to show species-specific response curves based on effort-corrected data.

To test if we could compensate for the absence of effort information in the opportunistic whaling data set, we derived a relative index of SPUE using a proportional sighting rate based on the fraction of total sightings in each cell that consisted of the specific species in question. We generated and compared proportional and standard encounter rates for dedicated IWC-IDCR survey data for a number of species. Both types of encounter rate were significantly and positively correlated for most species (e.g. p < 0.0001, Spearman’s rho = 0.88 for minke whales). These results indicated that the developed effort proxy might indeed represent a good approximation of SPUE or CPUE (catch per unit effort) for data sets with missing effort information if multiple species were surveyed simultaneously. Based on the assumption that whalers would have caught any species of whale where and whenever they encountered it, we subsequently computed proportional catch rates for individual species for each cell using the whaling data set and were thus able to
Table 3. Habitat usage in terms of depth, mean annual SST, and distance to the edge of sea ice for selected marine mammal species. Superscripts denote the particular habitat type about which the reference provided information: depth usage, temperature usage, and distance to edge of sea ice, respectively. For species marked by asterisk, distance from land was used as an additional constraining factor, limiting species to waters <500 km (*) from land (cont.: continental; estuar.: estuarine; reg.: regularly; pref.: preferably; assoc.: association; max.: maximum; Med: Mediterranean Sea; Black S.: Black Sea).

<table>
<thead>
<tr>
<th>Common name</th>
<th>Depth range</th>
<th>Temperature range</th>
<th>Distance to ice edge</th>
<th>General area</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clymene dolphin</td>
<td>Mainly coastal–continental</td>
<td>Full-on tropical</td>
<td>No assoc. with ice-edge</td>
<td>N Pacific – (Lat: &lt;10°N)</td>
<td>Davis et al. (1999), Reijnders et al. (1993), Rice (1998), Smith et al. (1986)</td>
</tr>
</tbody>
</table>

Note: The table continues with similar information for other species, detailing their preferred habitats, temperature ranges, and distances from the ice edge, with references provided for each entry.
Table 3 (continued)

<table>
<thead>
<tr>
<th>Common name</th>
<th>Depth range</th>
<th>Temperature range</th>
<th>Distance to ice edge range</th>
<th>General area minus (excluded areas)</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mediterranean monk seal</td>
<td>Mainly coastal–upper cont. shell to upper cont. slope</td>
<td>Subtropical only</td>
<td>No assoc. with ice edge, nowhere near ice at any time of the year</td>
<td>N hemisphere – (Indian Ocean, Pacific, Lon: &gt;20°W)</td>
<td>Duguy (1975) a, Kenyon (1981) a, Reijnders et al. (1993) b, c</td>
</tr>
</tbody>
</table>
generate effort-corrected response curves of opportunistic whaling data.

Finally, we compared the 3 types of large-scale response curves for all 5 species and all predictors to assess impact of effort biases and to evaluate our choice of assigned niche categories and the generic trapezoidal niche category shape itself.

**Model evaluation—RES model outputs.** We evaluated the generated RES predictions by testing the extent to which these may describe the variations in actual species’ occurrence for a number of marine mammal species found in different parts of the world’s oceans using sightings and catch data collected during dedicated surveys. Species for which we tested predictions were harbor porpoises *Phocoena phocoena*, northern fur seals *Callorhinus ursinus*, killer whales *Orcinus orca*, southern bottlenose dolphins *Lagenorhynchus cruciger*, sperm whales, blue whales, fin whales, humpback whales, dwarf minke whales, and Antarctic minke whales. We selected species to cover a wide taxonomic, geographic, and ecological range to test the robustness of the generic RES approach. In addition, we chose test data sets that varied widely in geographic and temporal scope to assess at which temporal or spatial scale RES predictions may prove to be insufficient in capturing patterns of species’ occurrences. To minimize risks of circularity, we tried to ascertain that test data had not been used to contribute directly or indirectly towards any of the studies or species reviews used to select input parameter settings. Test data sets included: (1) the SCANS (small cetaceans in the European Atlantic and North Sea) data collected during a dedicated line-transect survey in the North Sea and adjacent waters in the summer of 1994 (Hammond et al. 2002), (2) a long-term catch/sighting data set of northern fur seals collected during annual dedicated sampling surveys in the northeastern Pacific that were conducted in collaboration by the United States and Canadian federal fisheries agencies (Department of Fisheries and Oceans [DFO]—Arctic Unit & National Marine Fisheries Service [NMFS]) between 1958 and 1974, and (3) the long-term IWC-DESS data set described above (IWC 2001b) (Table 4).

Standard evaluation approaches for habitat suitability models based on confusion matrices are greatly impacted by difficulties to distinguish between true absences of species from an area and apparent absences due to detectability issues or insufficient sampling effort (Boyce et al. 2002). We therefore developed an approach similar one recommended by Boyce et al. (2002) to test predictions of presence-only models. Specifically, we compared the predicted gradient in RES scores across all cells covered by a survey with an observed gradient of relative usage by a given species in these cells, as described by the encounter rates of a species during the surveys. Again, species-specific encounter rates were obtained by binning records from each data set by raster cells, using only those records with sufficient spatial and taxonomic accuracy (i.e. catch or sighting positions of reliably identified species were reported to, at least, the nearest half degree latitude/longitude). For the reasons described above, we used the minke whale sightings in the IWC-DESS database to test the predictions for both the Antarctic minke whale and the dwarf minke whale.

---

**Table 4. Sighting and catch data sets used for RES model testing (abbreviations for data sets and institutions see ‘Model evaluation—RES model outputs’)**

<table>
<thead>
<tr>
<th>Agency/Source</th>
<th>IWC-BIWS catch data</th>
<th>IWC-IDCR/SOWER survey data</th>
<th>SCANS survey data</th>
<th>Northern fur seal survey data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey area</td>
<td>World</td>
<td>Antarctica (south of 60°S)</td>
<td>greater North Sea</td>
<td>NE Pacific</td>
</tr>
<tr>
<td>Survey focal species</td>
<td>Large whales</td>
<td>Minke whales</td>
<td>Harporkpooise</td>
<td>Northern fur seal</td>
</tr>
<tr>
<td>No. of marine mammal species reported</td>
<td>~20</td>
<td>~50</td>
<td>~5</td>
<td>1</td>
</tr>
<tr>
<td>No. of sighting/ catch records</td>
<td>~2,000,000</td>
<td>~35,000</td>
<td>1940</td>
<td>~18,000</td>
</tr>
</tbody>
</table>
Using only ship-based sightings, species-specific
SPUEs were generated for the SCANS data set in the
same fashion used for the IWC-DESS data. However,
actual transect information was unavailable for the
northern fur seal data set, although it contained ab-
sence records. Consequently, a proportional SPUE
per raster cell was generated based on an approach
similar to that applied to the IWC whaling data (i.e. we
assumed that, on average, the total number of survey
records [absence and presence] reported for 1 cell was
representative of the effort spent surveying a cell).
For each test data set, we compared species-specific
SPUEs with the corresponding RES model output for
that species by averaging encounter rates over all cells
covered by any effort that fell into a specific RES class.
Using a bootstrap simulation routine, we generated
1000 random data sets, similar in terms of means,
ranges, and distribution shapes to the predicted data
set. We then used Spearman’s non-parametric rank
correlation test (Zar 1996, JMP 2000) to compare aver-
age observed encounter rates with corresponding RES
classes based on model predictions and randomly gen-
erated data sets. To assess the performance of our
model compared to random distributions, we obtained
a simulated p-value by recording the number of times
the relationship between random data sets and ob-
served SPUEs was as strong as or stronger than that
found between the observed encounter rates and our
model predictions.

RESULTS
Relative environmental suitability predictions

Using available expert knowledge, RES modeling
allows the prediction of potential distribution and habi-
tat usage on very large-scales across a wide range of
species in a standardized, quantitative manner. Model
results represent specific, testable hypotheses about
maximum range extents and typical occurrence pat-
terns throughout a species’ range averaged over the
course of a whole year at any time from 1950 to 2000.
Examples of RES predictions for 11 pinniped, 6 toothed,
and 3 baleen whale species are shown in Fig. 3A–C.
Examples were selected to demonstrate the applicability of the modeling approach over a wide
geographic and taxonomic range of species (com-
pare Table 1, present paper, with Kaschner 2004, her
Appendix 1) and to illustrate the diversity of generated
model outputs for species occupying different en-
vironmental niches. Where they existed, we included
published outlines of maximum range extents (e.g.
Jefferson et al. 1993, Reijnders et al. 1993) for com-
parison. RES predictions for all other species can be
viewed on-line at www.seaaroundus.org/distribution/
search.aspx and are available in Kaschner (2004).

Generally, maximum extents of RES predictions for
species closely matched published distributional out-
lines (Fig. 3). RES maps for many species also captured
distinct areas of known non-occurrence well, without
the need to introduce any geographic constraints.
Examples of these are the predicted absence of hooded
seals from Hudson Bay, the restriction of gray whales
to the NE Bering Sea, and the non-occurrence of
Irrawaddy dolphins in southern Australia.

RES modeling illustrates the degree of possible spa-
tial niche partitioning that is already achievable based
on the few basic environmental parameters. The com-
plexity of the relationships between these parameters
alone can lead to distinctly different patterns of suit-
able habitat for species with slightly different habitat
usages, such as those demonstrated by the predictions
for hooded seals (Fig. 3) and harp seal Pagophilus
groenlandica in the North Atlantic (Kaschner 2004).

Published maximum range extents of the 2 species,
which are similar in terms of size and diets (Reijnders
et al. 1993), suggest largely sympatric occurrences and
a high degree of interspecific competition. However,
small divergences in habitat usage of the 2 species
(Table 3, present paper, and Kaschner 2004) resulted
in predictions that suggest substantial spatial niche
separation and highlight the importance of habitat
preferences as a mechanism to reduce competition.

Model evaluation

Evaluation of species response curves and impacts of
effort biases

Results from the analysis of whaling data highlighted
the potential problems of using opportunistic data
in presence-only models on very large scales in the
marine environment. At the same time, results pro-
vided basic support for our selected niche category
shape and the use of published information to assign
species to niche categories.

Comparison of the distribution of catch ‘presence’
cells by environmental strata with globally available
habitat indicated that even quasi-cosmopolitan and
long-term opportunistic data sets such as the whaling
data may not be a representative sub-sample of the
habitat used by species with global range extents
(Fig. 4A,B). Most existing presence-only models gen-
erate predictions based on the investigation of the
frequency distribution of so-called presence cells in
relation to environmental correlates. However, our
analysis showed that simple species-specific catch
‘presence’ histograms that ignore the effects of hetero-
Fig. 3. Examples of RES model outputs: predicted RES (ranging from less suitable [light] to very suitable [dark]) based on habitat usage information for (A) 11 pinniped, (B) 6 odontocete and (C) 3 mysticete species. Outlines of proposed maximum range extent (Jefferson et al. 1993) are included for comparison. Note that, when viewed on a global scale, RES predictions for many coastal species are difficult to see in narrower shelf areas such as along the western coast of South America and eastern coast of Africa, and apparent absences from certain areas may just be artefacts of viewing scale. RES predictions of narwhal distribution in the Sea of Okhotsk are masked to some extent by those for the northern right whale dolphin. Similarly, predictions for New Zealand fur seals in Australia are masked by those for Australian sea lions. RES maps for all marine mammal species can be viewed on-line at www.seaaroundus.org/distribution/search.apx and are available in Kaschner (2004).

Fig. 4. Frequency distributions of: (A) globally available habitat and (B) amount of habitat covered by whaling effort as the percent of cells per available environmental stratum for depth, mean annual SST, and mean annual distance to ice edge.
geneously distributed sampling effort generally diverged substantially from bar plots of encounter rates obtained from dedicated survey data collected in the same area for all species investigated (see examples shown in Fig. 5A,C). In contrast, effort-corrected proportional catch rates by environmental strata closely resembled bar plots generated from dedicated surveys (Fig. 5B,C). Overall, all available information suggested that the trapezoidal shape of niche categories used in this model may be a reasonable approximation of marine mammal response curves for those species for which habitat usage could be investigated on larger scales.

In terms of depth ranges used, we generally observed a good fit between the niche categories we had assigned species to and the bar plots based on proportional catch rates and SPUEs, though not with those based on frequency distributions of catch ‘presence’ cells (Fig. 5). In contrast, with respect to temperature and distance to ice, we found great discrepancies between general current knowledge about the global habitat usage of many species and the respective species’ habitat use that was suggested by all bar plots for these 2 predictors (not shown). These findings suggested that predictions of global, year-round distributions generated by standard presence-only modeling techniques and based on the whaling data alone might not reflect total distributional ranges of these species well.

Evaluation of RES predictions

RES modeling captured a significant amount of the variability in observed species’ occurrences — corrected for effort—in all test cases (Table 5). Average species’ encounter rates were positively correlated with predicted suitability of the environment for each species, except for...
the dwarf minke whale (Table 5). For this species, RES predictions were significantly but negatively correlated with the generic minke whale records in the IWC-IDCR data set. In contrast, <1% of the random data sets produced results that were more strongly correlated with observed encounter rates than RES predictions in most cases (Table 5). Killer whales and blue whales were the only 2 species for which a higher percentage of random data sets showed an equally strong correlation with the observed SPUEs. Only for these 2 species chance cannot be excluded as a factor to explain the significance of the relationship detected between RES predictions and observed patterns of occurrence. Model predictions were fairly robust across a large range of temporal and spatial scales, as significant correlations were found even in the case of harbor porpoise using the comparatively small-scale and short-term SCANS data set.

**DISCUSSION**

RES predictions

Our model represents a new objective approach for mapping large-scale distributions of marine species using non-point data. Predictions represent the visualization of current expert knowledge about species occurrence with respect to some aspects of environmental heterogeneity that indirectly determine distribution boundaries and patterns of occurrence of species within these boundaries. RES model performance is convincing when compared to existing information about species’ distributions, available in the form of descriptions of occurrences (see e.g. Rice 1998), or existing sketched outlines of distributional ranges (Jefferson et al. 1993). RES predictions are based on clearly defined assumptions and parameter settings and are thus reproducible and testable—unlike sketched distribution maps that may vary considerably between sources owing to differences in underlying assumptions or subjective and possibly arbitrary decisions made by the expert who drew them. In addition, by sacrificing ‘detail for generality’ (Levins 1966, Gaston 1994) and utilizing non-point data such as expert knowledge, the RES model can accommodate the frequently poor quality of available species’ occurrence data that often precludes the use of other statistical habitat prediction approaches. Because our more process-orientated approach is based on information about a species’ general occurrence in ecological space, like other niche models, it may be applied beyond existing survey ranges in geographic space (Hirzel et al. 2002). Thus, RES modeling represents a useful tool to investigate different hypotheses about large-scale distributions over a broad range of species, including those for which only few sighting records exist. In summary, the principle strength of the RES model lies in its greater objectivity in comparison to hand-drawn range extent and its generic applicability and its ability to utilize non-point data in comparison to statistical habitat suitability models.

In most cases, the predicted relative environmental suitability corresponded closely to the present ecological niche of a species. In other cases, predictions approximated a species’ habitat, including its historical range extension prior to human-induced depletion. For
some species, however, our results diverge substantially from known distributional ranges, suggesting that other factors may play a more important role in determining distributions. In general, RES predictions should be viewed as hypotheses about major aspects of a species’ fundamental spatial niche.

**RES predictions: limitations and biases**

The predictions generated by our model are affected by various biases, operating at different levels. Some biases are inherent to the present implementation of our approach, such as the lack of consideration of other factors known to influence species’ occurrence or the definition and shape of species response curves or the model algorithm. Other biases are directly associated with the data sets used for dependent and independent variables.

**Other factors influencing species’ occurrence**

In most cases, the realized niche of a species is likely to be influenced by far more factors other than the 3 basic environmental parameters considered in our model, though the role these play will differ among species. Investigations of environmental correlates of species’ occurrence have identified a host of other parameters, such as warm core rings for sperm whales (Jaquet & Whitehead 1996), zones of confluence of cyclone–anticyclone eddy pairs for a number of cetacean species (Griffin 1999, Davis et al. 2002), or the depth of the bottom mixed layer for North Atlantic right whales (Baumgartner et al. 2003). Consequently, it can be expected that the incorporation of factors such as these would lead to more heterogeneous patterns of species’ occurrence than implied by our model results.

Dynamic ecological factors, such as intra- and interspecific competition and other behavioral interactions, also greatly influence the occurrence of species, especially on smaller geographic and temporal scales (Austin 2002). Such factors may considerably reduce niche overlap between different species as, for example, in the cases of Australian sea lions and New Zealand fur seals. These 2 species co-occur along the southern Australian coastline as implied by RES predictions (Fig. 4), but in reality occupy different niches within this region due to behavioral differences (Ling 1992).

RES modeling currently also ignores effects of seasonality and environmental regime shifts, as well as changes in habitat preferences or usage associated with different phases in the annual life cycle of a species. The lack of consideration of short-term and long-term temporal variation of environmental parameters will be most noticeable in areas with great interannual or seasonal fluctuations, such as for some areas along the east coast of the United States (Angel 1992, NOAA/NODC 1998) or during environmental regime shifts such as El Niño events. Likewise, discrepancies between known occurrences and RES predictions will be more pronounced for species undergoing extensive annual migrations or for those species with large increases or decreases in population size. Changes in habitat usages, well documented for many of the baleen whales (Kasuya & Miyashita 1997), often accompany the seasonal shifts from feeding to breeding grounds. Here, parameters other than those determining food availability may become important, such as predator avoidance (Corkeron & Connor 1999, Pitman et al. 2001). Similarly, it has been proposed that extreme fluctuations in population size and associated range depletions or expansions may result in changes in habitat usages over long temporal scales, especially in highly depleted, long-lived species such as the North Pacific right whale *Eubalaena japonicus* (Tynan et al. 2001).

Some of the most obvious discrepancies between RES predictions and known regional occurrences of species, however, may be explained by range depletions caused by past or present anthropogenic impacts, such as whaling, sealing, or fisheries bycatch. An example of the importance of this human-related factor is the stark contrast between the predicted distribution of the North Atlantic right whales (Fig. 3), and today’s well-known absence of this species from northeastern Atlantic waters (Perry et al. 1999), due to exploitation by whalers in past centuries (Brownell et al. 1983).

**Model algorithm biases**

Observed discrepancies between RES predictions and known species’ occurrences may also be due to biases inherent in the RES model algorithm and the assumptions about niche category shape and types, all of which are likely simplistic. A linear relationship between all 3 environmental parameters is improbable, as is the assumption that each of them will play an equally important role in influencing distributions across all species (as implied by our unweighted resource selection function). Likewise, the unimodal shape of niche categories—although found to be the most common type of functional responses in 1 terrestrial study (Oksanen & Minchin 2002) and to some extent supported by the investigation of large-scale species’ response curves conducted here (Fig. 5)—is unlikely to adequately describe the presences of mammal species along environmental gradients in marine
ecosystems in many cases. Although functional responses are probably strongly bimodal for some migratory species, the trapezoidal shape we used may, nevertheless, represent the most parsimonious and broadly applicable choice for predicting general annual average distributions. Investigations of marine mammal occurrence along environmental gradients in the past have been mainly restricted to relatively small scales, generally only encompassing a sub-set of the species’ range (Baumgartner 1997, Kasamatsu et al. 2000, Cañadas et al. 2002). In the future, a meta-analytical investigation of large-scale response curves for some of the more data-rich marine mammal species would allow us to improve our current assumptions and is therefore regarded as a high priority.

Our quantitative definitions of niche categories currently ignore geographical differences in factors that determine niche boundaries or community transition zones. For instance, in comparison to other parts of the world, the edge of the continental shelf is typically much deeper (~500 m) in Antarctic waters, where the weight of the ice has caused the continental plate to sink (Knox 1994a). Consequently, the definition of 200 m bottom depth as a cut-off point for shelf-edge categories (Table 2) resulted in predicted absences of many species in some Antarctic regions where these species are known to occur regularly in high numbers (Hedley et al. 1999, IWC 2001b).

Biases of dependent and independent variables

The environmental parameters used as predictors in our model were affected by biases, which include direct measurement errors associated with the samples, and problems introduced through interpolation and rasterization processes (for detailed reviews of biases please refer to data providers, such as http://nsidc.org/data/ssmr_ssmi_ancillary/trends.html#gis and NOAA/NODC 1998). Long-term averages of SST measurements will have been particularly affected by interpolation issues due to the temporally heterogeneous sampling effort over the past 50 yr (NOAA/ NODC 1998). As a result, RES predictions may be biased towards time periods of higher sampling effort. Long-term ice edge data is affected by similar biases, but RES predictions were also influenced by the manual smoothing of ice edges, undertaken to eliminate nonsensical results in the computation of ice edge distances. In some cases, this smoothing resulted in predicted false absences or presences of species, such as the absence of harbor porpoise from the Baltic and Sea of Azov (Kaschner 2004). Furthermore, predictions were affected by the use of simple presence/absence ice data which did not allow the distinction between fast-ice (e.g. Weddell seals Leptonychotes weddellii; Kaschner 2004) and pack-ice species (e.g. Ross seal, Fig. 3). In the future, some ice data biases may be reduced by the use of more flexible sea ice concentration gradients instead of rigid presence/absence thresholds.

Unlike the independent variables, the information forming the basis for our dependent variables is less likely to be affected by interpolation issues, due to its mainly qualitative nature. Nevertheless, skewed effort distribution is likely to have had some influence on the current general perceptions about the habitat usage of many species.

Model evaluation

Evaluation of species response curves and impacts of effort biases

Investigation of the relationships between a species' occurrence and existing environmental gradients— which forms the basis of all habitat suitability models— requires adequate coverage of the habitat available to this species both in space and time (Manly et al. 2002). Comparison of the proportion of habitat covered by whaling operations with globally available habitat suggested that, even for very large opportunistic data sets such as the whaling data, sampling effort might not be equally representative of all habitat that is available to species with known cosmopolitan distributions. Though unbiased sampling effort is a key assumption also for presence-only models (Hirzel & Guisan 2002), predictions of terrestrial species’ distributions generated by GARP, for instance, have been shown to be relatively insensitive to heterogeneously distributed effort (Peterson 2001, Stockwell & Peterson 2001). However, in comparison to terrestrial systems, insufficient coverage of available habitat due to spatially and temporally skewed effort is likely much more pronounced in the marine environment, where weather conditions and sheer distances restrict survey efforts mainly to the summer months and to areas relatively close to ports.

The importance of effort considerations was illustrated by the comparison of species’ response curves to environmental gradients based on opportunistic data sets and those derived from effort-corrected data or available habitat usage information. Minke whales, for instance, are generally perceived to be closely associated with coastal and shelf waters (Jefferson et al. 1993)— a perception which is supported by statistical investigations of minke whale occurrences in relation to depth throughout the world (Sigurjónsson 1995, Kasamatsu et al. 2000, Hamazaki 2002, Moore et al. 2002) and is reflected by our choice of niche category.
However, this perception was greatly at odds with the depth distribution of minke whale catches in the whaling data, which—even if catch numbers were corrected for proportionally available habitat—suggested a predominant usage of much deeper waters for this species. The high number of minke whale catches reported in offshore areas might be explained by the concentration of whaling activities in deeper waters, where the larger whale species, such as blue, fin, and sperm whales that initially represented the primary targets of whalers, were predominately known to occur (Perry et al. 1999). Minke whales did not become a target species until quite late in the whaling era, but were likely nevertheless caught on a regular basis if whalers happened upon them. The sheer amount of whaling effort in deeper waters thus masked this species’ actual habitat usage if analyses were based on frequency of catch ‘presence’ cells alone.

In contrast, bar plots of effort-corrected catches were consistent with the general perception of depth usage of minke whales. The use of proportional encounter rates to investigate species’ response curves might therefore help to compensate for some effort biases. In combination with results from other studies of cetacean response curves (e.g. Kasamatsu et al. 2000, Cañadas et al. 2002), bar plots of encounter rates based on both whaling data and dedicated surveys provided good support for the trapezoidal shape of niche categories used here.

In our analysis, we chose to ignore all temporal aspects of the data sets. The binning of catches across all years will have masked effects of the well-known serial depletion of the large whale species (Clark & Lamberson 1982, Perry et al. 1999) and the distortion likely introduced by any progressive spatial expansion of catch effort (Walters 2003). In view of these temporal biases and the very different time periods during which whaling data and the IWC-DESS survey data were collected, the similarity of encounter rate bar plots based on the 2 data sets was quite surprising. We propose that these findings provide indications that general usage of habitat by the species investigated here may have been quite consistent over the last century, despite the considerable fluctuating in population sizes.

The extent to which species’ response curves from opportunistic data sets may be representative of habitat usage throughout a species’ range appears to depend on the type of environmental predictor. The good fit of encounter rate bar plots and selected niche category in terms of bottom depth across almost all species indicated that whaling records indeed reflect the predominant perception of a species’ global depth usage—if effort is taken into consideration. However, comparison of general current knowledge about global habitat usage in terms of temperature and ice distance—as represented by our selected niche categories for the different species—with bar plots for these 2 predictors suggested that catch data distributions were strongly seasonally biased. Whaling effort was concentrated in the polar waters of both hemispheres during summer months (IWC 2001a), thereby only covering parts of the distributions of most species targeted, namely their summer feeding grounds. While a species’ depth preference is often consistent throughout its latitudinal range extent, temperature ranges and distance to ice edge will tend to vary depending on when and where throughout its range and annual life cycle an animal is captured or sighted. Thus, from the perspective of modeling highly migratory species with global distributions in the marine environment, reliance on available point data sets alone would likely result in a biased prediction, despite the potentially broad geographic coverage and large sample sizes of such data sets. In contrast, RES outputs may represent more balanced predictions of annual average distributions of cosmopolitan or quasicosmopolitan species, since we were able to supplement seasonally biased point data with additional sources of information about general occurrences during other times of the year when we assigned species to specific niche categories (Table 4).

In conclusion, our analyses of whaling data suggested that for habitat prediction on very large scales it may be difficult to find data sets that would allow the straightforward application of presence-only habitat suitability models. Nevertheless, a quantitative comparison of the quality of RES predictions for quasicosmopolitan marine species with those generated by other niche models using available opportunistic data sets is needed to allow a more rigorous investigation of the effects of skewed effort distributions on very large scales.

**Evaluation of RES predictions**

Statistical tests of RES model results indicated that our generic approach has some merit to adequately describe suitable habitat, as significant amounts of the variability in average species’ occurrence were captured for all but 1 species tested (Table 5). In contrast, simulated random data sets rarely showed equally strong or stronger relationships with the observed data. Several factors may explain the 2 cases in which random data sets often showed equally strong relationships with the observed data. For blue whales, the observed number of encounters was very low, possibly leading to the relatively weak correlation between
predicted RES values and the test data set (Table 5). For killer whales, several different ecotypes or subspecies occupy distinctly different ecological niches in different parts of the world, including Antarctic waters (Pitman & Ensor 2003). To capture the preferred habitat of all subspecies, we selected very broad niche categories. Likewise, the IWC-IDCR data set does not distinguish between different subspecies, as these are difficult to identify in the wild. The very broad predictions and the mixed sightings pool of subspecies with different habitat usage may have contributed to the large proportion of random data sets that could explain the observed variation in the test data set equally well. Similarly to the mixed pool of killer whale sightings, the generic ‘minke whale’ observations in the test data set likely represent sightings of both the Antarctic minke whale and the dwarf minke whale—2 species which appear to prefer slightly different habitats (IWC 2001b, Perrin & Brownell 2002, Matsuoka et al. 2003). Interestingly, RES predictions for the Antarctic minke whale were positively correlated with the generic sightings, while our predictions for the sister species showed an equally significant but negative correlation with these sightings. This suggests that either all minke whales encountered in the survey belonged to just 1 species, the Antarctic minke whale, or—and this is more likely—our model exaggerated the niche separation between the 2 species.

**Independence of test data**

The statistical testing of both our predictions and model assumptions are affected by a number of biases. First, given the broad nature of our niche categories and the type of information they were based on, we cannot be certain that the test data sets were indeed completely independent. Consequently, there is a risk of circularity, if the test data had somehow formed the basis of one of the broad ‘expert knowledge’ statements (such as ‘coastal’ and ‘subtropical’ species) that was fed into our model. However, the process of abstraction from point data to these general statements, in and of itself, would probably ensure a certain degree of data independence. Furthermore, we argue that—even if test data did serve as a basis of niche descriptions—testing the extent to which such broad statements may actually suffice to describe species’ presences and absences when applied in a GIS modeling framework is a worthwhile exercise. Nevertheless, we tried to minimize potential circularity by excluding all references that were directly based on these data from our pool of input sources used to determine niche settings for the particular species tested (e.g. Kasamatsu et al. 2000, Hammond et al. 2002).

**Comparison with other habitat suitability modeling approaches**

The validation analysis indicated a remarkable robustness of RES predictions across a broad range of temporal and spatial scales and for a wide taxonomic range of species, suggesting that species’ distributions and patterns of occurrence in the marine environment may be quantitatively described using surprisingly few basic parameters. Despite the apparent robustness of the RES modeling approach to perform well at different scales, care should be taken when interpreting model outputs.

It is highly unlikely that our more mechanistic model will be capable of predicting the real probability of species’ occurrences in a specific place on a specific day or month of a given year. The RES model should therefore not be viewed as an alternative to empirical presence/absence type habitat prediction approaches that can and should be applied on smaller geographic scales to predict marine mammal occurrence when and where dedicated line-transect data sets are available. Similarly, the application of more sophisticated presence-only models, such as GARP or ENFA, may often be preferable at intermediate scales and when available data sets can be shown to represent a geographically and temporally unbiased subsample of the habitat available to a species. However, there is some indication—based on the analysis of whaling data—that effort biases might be more prominent in the marine environment than in terrestrial systems, thus potentially precluding the straightforward use of available opportunistic point data sets in presence-only models, though this remains to be investigated in more detail. In general, the quality of predictions generated by any model can only be as good as the available data, and more sophisticated models do not necessarily perform better than simpler approaches, especially if data quality is poor (Moisen & Frescino 2002). Consequently, RES modeling may be more suitable than other niche models on very large scales, where available data sets may not be representative of the species’ actual occurrence or if point data are completely missing.

**Future work and applications**

In the future, RES modeling may serve as a useful tool to address both basic ecological questions as well as management and conservation-related issues in situations where the paucity of comprehensive point data sets—a situation commonly encountered in the marine environment—precludes the use of other more data-intensive habitat modeling approaches. Relying on
more readily available types of data, such as expert knowledge, RES modeling will be particularly useful to study basic niche similarities and overlap between different marine species or groups of species on very large scales. Its application may also be a worthwhile first step in investigating scientific questions challenged by the paucity or complete lack of existing occurrence records, including historical distributions of heavily depleted species (e.g. gray whales in the North Atlantic; Mitchell & Mead 1977), calving grounds of endangered baleen whale species (yet unknown for species such as the North Pacific right whale; Gaskin 1991), or changes in species distributions due to environmental regime shifts or climate change (K. Kaschner unpubl. data).

Most importantly, however, the extent to which RES-generated hypotheses describe observed patterns in species’ occurrence will allow more process-orientated questions to be asked about the role that other factors play in determining actual distributions. Similarly, the quantitative comparision of RES predictions with other niche models, such as GARP or ENFA, will help identify discrepancies that may be symptomatic for underlying sampling biases and related issues. This may help to highlight the problems of skewed effort distributions for habitat suitability modeling in the marine environment on very large scales. Future evaluation of RES predictions for species with available sighting data sets using standard evaluation statistics based on confusion matrices and thresholds optimized by receiver–operator curves for species presence would be helpful for a case-by-case investigation of the extent to which our predictions correspond more closely to a species’ fundamental versus its realized niche.

In a management context, RES predictions can usefully supplement small-scale studies by providing some greater context of general boundaries and potential focal areas of species’ occurrences in unsurveyed regions. Thus, the RES model may provide cost-efficient starting points to focus future research and survey efforts. This is especially practical when dealing with the many data-poor species in the lesser-studied regions of the world, such as some of the rare and endangered beaked whales. The usefulness of habitat prediction models to minimize anthropogenic impacts on endangered species of marine mammals through the implementation of effectively designed marine reserves has already been demonstrated on relatively small scales (Mullin et al. 1994b, Moses & Finn 1997, Hooker et al. 1999). RES modeling may be equally useful when attempting to delineate efficient marine protected areas or critical habitat on larger geographic scales, by generating global spatialy explicit indexes of biodiversity and species richness, or visualizing potential geographic hotspots of high conflict with fisheries or other human operations (Kaschner 2004, K. Kaschner et al. unpubl. data).

Acknowledgements. This research was conducted as part of the ‘Sea Around Us’ project, initiated and funded by the Pew Charitable Trusts of Philadelphia, USA, and with support from the National Science and Engineering Research Council of Canada. A ‘Li Tze Fong’ graduate fellowship and a partial university graduate fellowship from the University of British Columbia provided additional support for K. Kaschner. We are grateful to K. Bodtger, E. Gregr, A. Hall, and B. Wilson (MMRU-UBC), as well as S. Jansen (BFA-Hamburg) for many fruitful discussions on methodology and model development. K. Willis helped compile habitat usage information, and M. Denoth, A. Winship, and R. Joy offered invaluable advice and assistance during the development of the validation approach. Comments from Falk Huettmann and the anonymous reviewers greatly helped to improve this manuscript. We are indebted to C. Allison and the Secretariat of the International Whaling Commission for providing us with the IWC catch and sighting data sets. We also thank P. Hammond, SMRU, for the SCANS data set and P. Olesiuk, DFO, for the northern fur seal data. In addition, we are grateful for the ice edge data provided by the United States National Snow & Ice Data Centre.

LITERATURE CITED


Camps Casado, Sagarminaga R, García-Tiscar S (2002) Cetacean distribution related to depth and slope in the Mediter-


Kenney RD, Winn HE (1986) Cetacean high-use habitats of the northeast United States continental shelf. Fish Bull 84:345–357


Lalas C, Bradshaw CJA (2001) Folklore and chimerical numbers: review of a millennium of interaction between fur


MacLeod CD (2005) Niche partitioning, distribution and competition in North Atlantic beaked whales. PhD, University of Aberdeen, UK


Mörzer Bruyns WJF (1971) Field guide of whales and dolphins. Tor, Amsterdam


Kaschner et al.: RES mapping of marine mammal distributions 309


Editorial responsibility: Howard I. Browman (Associate Editor-in-Chief), Storebø, Norway

Submitted: March 18, 2005; Accepted: September 21, 2005

Proofs received from author(s): May 17, 2006