ECOGRAPHY

Research

Geographical differences in habitat relationships of cetaceans across an ocean basin

Laura Mannocci, Jason J. Roberts, Eric J. Pedersen and Patrick N. Halpin

L. Mannocci (https://orcid.org/0000-0001-8147-8644) ✉ (laura.mannocci@gmail.com), MARBEC, Univ. Montpellier, CNRS, Ifremer, IRD, Sète, France. – J. J. Roberts, P. N. Halpin and LM, Marine Geospatial Ecology Laboratory, Nicholas School of the Environment, Duke Univ., Durham, NC, USA. – E. J. Pedersen, Dept of Biology, Concordia Univ., Montreal, QC, Canada.

The distributions of highly mobile marine species such as cetaceans are increasingly modeled at basin scale by combining data from multiple regions. However, these basin-wide models often overlook geographical variations in species habitat relationships between regions. We tested for geographical variations in habitat relationships for a suite of cetacean taxa between the two sides of the North Atlantic basin. Using cetacean visual survey data and remote sensing data from the western and eastern basin in summer, we related the probability of presence of twelve cetacean taxa from three guilds to seafloor depth, sea surface temperature and primary productivity. In a generalized additive model framework, we fitted 1) basin-wide (BW) models, assuming a single global relationship, 2) region-specific intercepts (RI) models, assuming relationships with the same shape in both regions, but allowing a region-specific intercept and 3) region-specific shape (RS) models, assuming relationships with different shapes between regions. RS models mostly yielded significantly better fits than BW models, indicating cetacean occurrences were better modeled with region-specific than with global relationships. The better fits of RS models over RI models further provided statistical evidence for differences in the shapes of region-specific relationships. Baleen whales showed striking differences in both the shapes of relationships and their mean presence probabilities between regions. Deep diving whales and delphinoids showed contrasting relationships between regions with few exceptions (e.g. non-statistically different shapes of region-specific relationships for harbor porpoise and beaked whales with depth). Our findings stress the need to account for geographical differences in habitat relationships between regions when modeling species distributions from combined data at the basin scale. Our proposed hypotheses offer a roadmap for understanding why habitat relationships may geographically vary in cetaceans and other highly mobile marine species.

Keywords: environmental predictors, geographical variation, habitat relationships, highly mobile marine species, North Atlantic Ocean, species distribution modeling

www.ecography.org

Introduction

Species distribution models (SDMs) are the most widely used statistical tools to examine relationships between species distributions and environmental predictors (Elith and Leathwick 2009). One of the core assumption of SDMs is that relationships between environmental predictors and species occurrences are invariant over the modeling region (Dormann et al. 2012, Miller 2012). In terrestrial species such as birds and amphibians, large geographical variations in patterns of habitat associations have been found, with modeled relationships differing substantially in shape between regions (Whittingham et al. 2007, Zanini et al. 2009, Schaub et al. 2011). However, geographical variations in habitat relationships have been less studied in highly mobile marine species characterized by extensive distributional ranges such as seabirds, sea turtles, sharks and cetaceans. Such geographical variations in habitat relationships may be a reason for poor spatial transferability of SDMs in ecological and conservation applications (Sequeira et al. 2018), as found in some seabirds (Torres et al. 2015) and cetaceans (Redfern et al. 2017).

Marine survey programs implemented over the last decades have allowed collecting occurrence and abundance data on highly mobile species, but their coverages greatly vary between regions (Kaschner et al. 2012). In response to pressing marine resources management and conservation needs, a growing number of basin-wide SDMs have been developed by combining survey data from multiple regions (Mannocci et al. 2015, Cañadas et al. 2018, Virgili et al. 2019). In these basin-wide SDMs, global habitat relationships are modeled and thus, geographical variations in habitat relationships between regions are generally overlooked. In this study, we tested for geographical variations in habitat relationships for a suite of cetacean taxa using data from two separate regions of an ocean basin.

Satellite remote sensing generates environmental data over basin scales that are compatible with the extensive distributional ranges of cetaceans, providing readily available predictors for modeling these species distributions. Environmental predictors may describe important characteristics of the physical marine environment (e.g. sea surface temperature), or proxies of dynamic processes that enhance resource availability (e.g. primary productivity) (Redfern et al. 2006). While cetaceans are unlikely to respond to temperature or primary productivity directly, their prey, which are ectothermic and characterized by intermediate trophic levels, are expected to respond to these predictors more directly. Correlative SDMs based on statistical relationships between environmental predictors and cetacean occurrences still provide little insight on the mechanisms underlying cetacean interactions with their environment (Dormann et al. 2012, Palacios et al. 2013). Nevertheless, these relationships effectively describe cetacean habitats and underlie spatial predictions that are crucial for management and conservation applications given ever-increasing risks to these species from human activities and global change (Becker et al. 2012, Roberts et al. 2016, Mannocci et al. 2017, Redfern et al. 2017).

Using visual survey data collected in two regions located on either side of the North Atlantic basin, we tested for geographical variations in relationships between the probability of presence of twelve cetacean taxa and three broadly available environmental covariates (primary productivity, sea surface temperature and seafloor depth). Our statistical approach based on hierarchical generalized additive models effectively accounts for region-specific differences when developing SDMs of highly mobile marine species over basin scales.

Material and methods

Cetacean visual survey data

We relied on visual sightings of cetaceans from systematic aerial and shipboard line transect survey programs conducted in two regions of the western North Atlantic (WNA) and the eastern North Atlantic (ENA) in summer between 2004 and 2007 (Supplementary material Appendix 1). Surveys represented comparable effort on both sides of the basin (20 763 km in the WNA and 19 823 km in the ENA). All visual survey data were collected following a standard distance sampling methodology (Buckland et al. 2001). For each cetacean sighting, observers recorded identification to the closest possible taxonomic level (species, genus or family), group size, perpendicular distance and detectability covariates (e.g. sea state).

Twelve cetacean taxa (including ten species, one genus and one family) that were sighted on both sides of the North Atlantic basin were retained for the analysis. Taxa were grouped into three guilds (baleen whales: 340 sightings; deep diving whales: 445 sightings; delphinoids: 960 sightings) (Supplementary material Appendix 1 Table A1-3). Sighting distributions in both regions are shown for the three guilds in Fig. 1 and for individual taxa in Supplementary material Appendix 2.

For each taxon, effective strip half widths (ESHWs), giving the distances effectively surveyed on either side of the transects, were estimated by fitting detection functions to sightings recorded by observers using conventional and multiple covariate distance sampling (Buckland et al. 2001), and then ranking these functions according to Akaike's information criterion (AIC). To obtain sufficient sightings for detection function fitting, survey programs with similar observation protocols (e.g. using binoculars versus naked eye) were pooled (Roberts et al. 2016). All detection functions were fitted using R mrds package ver. 2.1.10 (Laake et al. 2014) and documented by Roberts et al. (2016) (results are reported in Supplementary material Appendix 1 Table A1-4).

Survey transects were projected to an Albers equal area map projection, and then divided into segments approximately 10 km long (Roberts et al. 2016). The effective area surveyed for each segment was calculated as twice the ESHW multiplied by the segment's length.
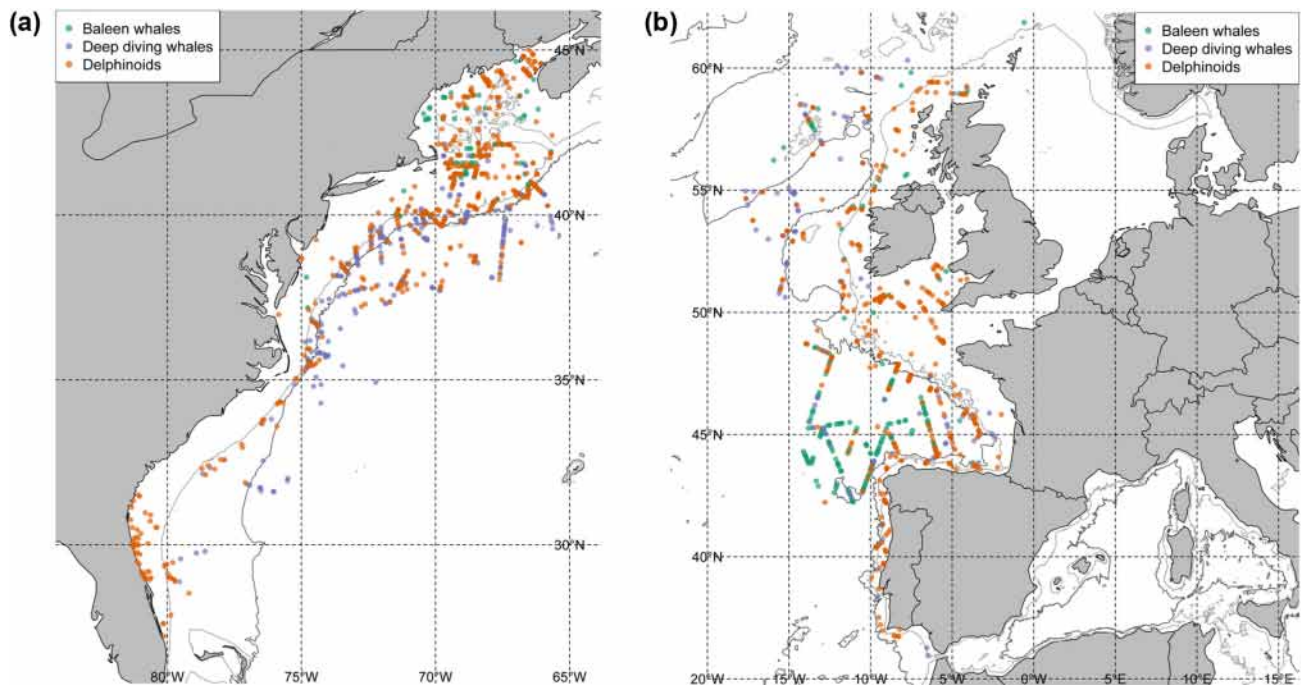
Figure 1. Maps of sightings of the three cetacean guilds from line transect surveys in the WNA (a) and ENA (b). The 200 m isobath in light grey represents the continental shelf break. The 2000 m isobath in dark grey represents the outer continental slope. Maps of sightings for individual taxa are shown in Supplementary material Appendix 2.

Environmental covariates

We considered three environmental covariates that are readily available over basin scales and known to affect the distribution of cetaceans or the distribution of their main prey species. These covariates are classical predictor variables in cetacean distribution modeling studies (Roberts et al. 2016, Tobeña et al. 2016, Lambert et al. 2017, Mannocci et al. 2017, Pennino et al. 2017, Rogan et al. 2017, Fiedler et al. 2018, Becker et al. 2019, Chavez-Rosales et al. 2019, Virgili et al. 2019). Seafloor depth, a static covariate, strongly affects the distribution of cetaceans (Cañadas et al. 2002, Yen et al. 2004). It was derived from SRTM30_PLUS global bathymetry available at a 30 arc second resolution (<https://topex.ucsd.edu/WWW_html/srtm30_plus.html>). Sea surface temperature (SST) shapes the distribution of ectothermic prey via physiological processes (Dolomatov et al. 2013). It was derived from the GHRSST Level 4 CMC 0.2deg Global Foundation Sea Surface Temperature Analysis (<https://podaac.jpl.nasa.gov/dataset/CMC0.2deg-CMC-L4-GLOB-v2.0>) at 1 day (d), 0.2° resolution. Primary productivity (PP) influences mid-trophic level prey of cetaceans (including zooplankton, fish and squid) via trophic coupling (Croll et al. 2005). PP was derived from the vertically generalized production model (Behrenfeld and Falkowski 1997) (<www.science.oregonstate.edu/ocean.productivity/>) at 8 d, 9 km resolution and was log-transformed to reduce skewness.

After obtaining the original covariate grids, we projected them to the 10 × 10 km resolution Albers equal area grid used for the analysis, then sampled them at the survey segment centroids. We used bilinear interpolation for depth and trilinear interpolation with the date of the segment as the time dimension for dynamic covariates. While the range of sampled depth values was similar between regions, the sampled ranges of SST and PP were broader in the WNA (which included some warmer and less productive waters not present in the ENA) (Supplementary material Appendix 3). We performed all geoprocessing using ArcGIS 10.2.2 and the Marine Geospatial Ecology Tools software (Roberts et al. 2010).

Statistical modeling

Nonlinear smooth relationships were estimated between each environmental covariate and the probability of presence of each cetacean taxon using three types of models in a generalized additive model (GAM) framework (Wood 2017): 1) basin-wide (BW) models that assumed a single global habitat relationship (one per covariate) in the combined regions, 2) region-specific intercept (RI) models that assumed habitat relationships with the same shape in both regions, but allowed the intercept to vary to account for differing mean presence probabilities, and 3) region-specific shape (RS) models that assumed habitat relationships with different shapes between regions.

In order to estimate relationships free from confounding statistical effects of potentially collinear covariates included the same model (Zuur et al. 2010), we only considered single-covariate binomial GAMs.

To account for varying cetacean detectability between segments, we used a complementary log–log link function and

the logarithm of the effective area surveyed as an offset (Kery and Royle 2015). We used the restricted maximum likelihood method for automated selection of smoothing parameters (Wood 2011). We set the basis dimension to 20 for the smoother in each model to allow for potentially complex nonlinear relationships. All models were fitted using R mgcv package ver. 1.8-26 (Wood 2018).

BW models, defined as follows, included a single global smoother with a thin plate regression spline basis with shrinkage, allowing covariates with non-significant relationships to be identified:

$$P_j \sim \text{Binomial}\big(\mu_j\big)$$

$$\log\Big(-\log\big(1-\mu_j\big)\Big) = b_0 + \log\big(A_j\big) + f\big(x_j\big)$$

where:

P_j is the observed presence/absence in each segment j, assumed to be binomially distributed
μ_j is the estimated probability of presence for segment j
b_0 is the intercept term
A_j is the effective segment area for j (model offset)
f is the smooth function of environmental covariate x
RI models were defined as BW models, but included region-specific intercepts:

$$\log\Big(-\log\big(1-\mu_j\big)\Big) = b_0 + b_r \times r_j + \log\big(A_j\big) + f\big(x_j\big)$$

where r is a factor variable representing the region (WNA or ENA) and b_r is the region-specific intercept.

RS models included a factor–smooth interaction basis (Wood 2017) that allowed relationship shapes to vary between regions but assumed similar complexity (i.e. a hierarchical GAM; Pedersen et al. 2019). RS models were defined as:

$$\log\Big(-\log\big(1-\mu_j\big)\Big) = b_0 + b_r \times r_j + \log\big(A_j\big) + f\big(x_j, r_j\big)$$

where f is the region-specific smooth of environmental covariate x.

To test whether cetacean occurrences were best modeled with global or region-specific relationships, we compared the fit of BW and RS models to the combined cetacean survey data on the two sides of the North Atlantic, using a version of the Akaike information criteria (AIC) (Burnham and Anderson 2002) designed to account for the effect of smoothing penalties on degrees of freedom (Wood et al. 2016). Then, to test whether the shape of relationships with the examined environmental covariates rather than just the mean cetacean presence probability differed between regions, we compared the fit of RS and RI models. A better fit of the RS model would indicate statistical evidence for differences in the shape of a given habitat relationship between regions. In contrast, a better fit of the RI model would indicate region-specific intercepts were sufficient to capture the difference in mean presence probabilities between regions. Finally, we visually assessed the differences in region-specific relationships estimated from the RS models.

For two species (sei whale and harbor porpoise), the majority of sightings occurred in a very limited range of environmental covariates values. This resulted in 'complete separation' when fitting models, a known issue in binomial regression (Gelman and Hill 2006). Complete separation occurs where it is possible for multiple combinations of parameters in GAMs to lead to similarly shaped functions, thus resulting in very large confidence intervals for several domains of covariate ranges (Supplementary material Appendix 4). The complete separation issue was accounted for by replacing all zeros in the cetacean visual survey data with a very small but non-zero value (set at 0.001). This was equivalent to assuming that even when the taxon was not observed in a given segment, it had a small but non-zero chance (0.1%) of being present. This approach stabilized confidence intervals without changing the average relationships between cetacean presence probabilities and covariates (Supplementary material Appendix 4 Fig. A4-1).

Results

Overall, RS models showed significantly better fits than BW models to cetacean survey data from the two sides of the North Atlantic (Table 1), indicating cetacean occurrences were best modeled with region-specific relationships. The RS model yielded significantly better fit (ΔAIC > 2) for 10 taxa with depth as the covariate and for 11 taxa with SST and with PP as the covariate (representing 32 of the 36 taxon-covariate instances, or 89%). Support for RS models over BW models was particularly strong for baleen whales, as indicated by the large ΔAIC.

RS models yielded significantly better fits than RI models in 23 of the 36 taxon-covariate instances (or 64%) (Table 1), indicating that for most taxa and covariates, the shapes of the relationships differed between regions, rather than just the mean probability of presence. RI models yielded significantly better fits in only 14% of the instances (for beaked whales and harbor porpoise with depth, minke whale and bottlenose dolphin with PP, and pilot whales with SST). This indicated that in these instances, relationship shapes were not significantly different between regions, and differences in mean presence probabilities between regions were effectively captured by region-specific intercepts.

A visual comparison of region-specific relationships further revealed ecologically important differences in habitat relationships of cetaceans between regions. Baleen whales showed striking differences in their relationships with all three environmental covariates between regions (Fig. 2), in accordance with their distinct distributions in the WNA and the

Table 1. Comparison of the fits of basin-wide (BW), region-specific shape (RS) and region-specific intercept (RI) models relating cetaceans' presence probability to seafloor depth, sea surface temperature (SST) and primary productivity (PP) (log-transformed), respectively. Comparison of the fits is indicated using change in AIC (ΔAIC). Colour coding for ΔAIC (BW versus RS): no colour when the RS model fits best and yellow when the BW model fits best. Colour coding for ΔAIC (RI versus RS): no colour when the RS model fits best and blue when the RI model fits best. The dash (–) character indicates that abs(ΔAIC) was ≤ 2 (undistinguishable difference in model fit). Detailed modeling results are available in Supplementary material Appendix 5.

Cetacean taxa	Depth		SST		PP	
	ΔAIC (BW versus RS)	ΔAIC (RI versus RS)	ΔAIC (BW versus RS)	ΔAIC (RI versus RS)	ΔAIC (BW versus RS)	ΔAIC (RI versus RS)
Baleen whales						
Fin whale	162	142	9	11	57	51
Minke whale	30	–	19	–	11	−3
Sei whale	28	29	14	15	29	26
Deep diving whales						
Beaked whales	–	−5	5	–	3	–
Sperm whale	6	5	5	6	6	5
Pilot whales	16	10	−8	−6	3	–
Risso's dolphin	4	3	25	15	17	–
Delphinoids						
Atlantic white-sided dolphin	6	8	5	6	8	8
Bottlenose dolphin	4	5	6	4	−7	−5
Short-beaked common dolphin	127	68	21	22	24	4
Striped dolphin	20	–	9	3	7	–
Harbor porpoise	−11	−10	7	8	16	14

ENA (found in on-shelf versus off-shelf waters, respectively; Fig. 1). For fin whale, both the shape of relationships and the average probability of presence differed between regions. For sei whale, relationships primarily differed in shape. For minke whale, the shape of relationships was comparable between regions but the average probability of presence was twice higher in the ENA relative to the WNA.

Deep diving whales showed contrasting relationships between regions (Fig. 2), except for pilot whales with SST, for which relationships decreased similarly and weakly in both regions. Beaked whales showed comparable depth optima between regions and a second higher depth optimum in the ENA only. Sperm whale and Risso's dolphin showed divergent shapes of region-specific relationships with SST and PP with overall higher probabilities of presence in the WNA relative to the ENA. Beaked whales and pilot whales showed more similar but weak region-specific relationships with dynamic covariates (with very low deviances explained; Supplementary material Appendix 5).

Delphinoids showed contrasting relationships between regions, except in a few situations (Fig. 2). The presence of harbor porpoise consistently peaked at very shallow depth in both regions. Bottlenose dolphin showed the same PP optimum over the commonly sampled PP range in the two regions. For Atlantic white-sided dolphin and bottlenose dolphin, depth relationships had similar shapes, but slightly steeper decreasing trends in the WNA. Contrasting region-specific relationships were found in all other situations, with particularly divergent shapes for harbor porpoise and short-beaked common dolphin with PP, striped dolphin with SST and short-beaked common dolphin with depth.

Discussion

As more studies address broad-scale marine resource management and conservation issues by developing basin-wide SDMs combining survey data from multiple regions (Mannocci et al. 2015, Cañadas et al. 2018, Virgili et al. 2019), it is becoming critical to examine whether the modeled habitat relationships vary between regions. We tested for geographical variations in habitat relationships for a suite of cetaceans across the North Atlantic basin. Overall, cetacean occurrences were better modeled with region-specific than with global habitat relationships, as revealed by the better fits of RS models over BW models (Table 1). The better fits of RS models over RI models for many taxa and covariates (Table 1) further provided statistical evidence for differences in the shapes of region-specific relationships. Our results stress the need to account for geographical differences in habitat relationships when modeling cetacean distributions from combined regional survey data. Our findings concur with those of Byrne et al. (2019) who found for another highly mobile marine predator – the shortfin mako shark – that relationships between shark resident behavior inferred from telemetry tracks and environmental covariates varied across ecosystems.

Seafloor depth, surface temperature and productivity have been effective predictors of diversity and distribution patterns of mobile marine species, including tuna, billfish and marine mammals at macroecological scales (Worm et al. 2005, Kaschner et al. 2006, Whitehead et al. 2008, 2010). Our results demonstrate that, while these predictors may relate to species occurrences in one way in one region, the shapes of relationships may drastically change in another region.
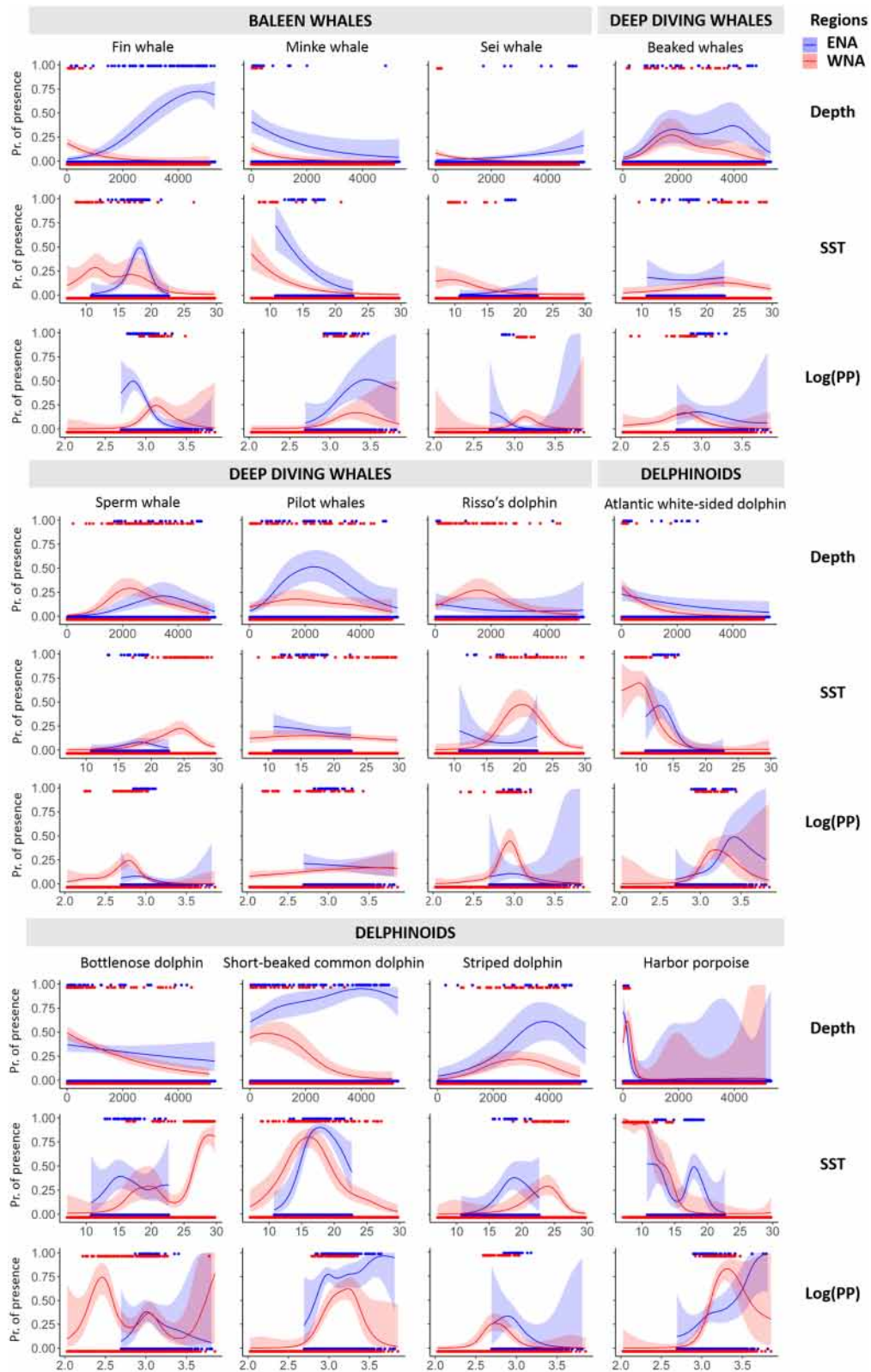
Figure 2. Relationships between cetaceans' presence probability and environmental covariates estimated from RS models. Shaded areas are 95% confidence intervals for each estimated relationship. Dots represent presence/absence data (1/0 probability of presence, respectively) for surveyed segments in the WNA and ENA (dots have been offset vertically for readability). The narrower SST and PP range in the ENA led to truncated relationships in the ENA. Relationships estimated from BW models are shown in Supplementary material Appendix 6.

The substantial differences in habitat relationships between the two sides of the North Atlantic suggests that the complex ecological rules governing cetacean distributions have yet to be elucidated.

We suggest some possible hypotheses for these detected geographical differences. Although these hypotheses are not testable with the current data, they represent a roadmap for future investigations that would help determine why habitat relationships may not be geographically consistent in cetaceans and other highly mobile marine species.

1) Different habitat uses between regions. Because marine predators show stronger habitat selection when foraging than transiting (Heerah et al. 2017), different relationships are expected between regions where their habitat uses differ. For sei whale for example, satellite telemetry and simultaneous prey studies suggest a summer feeding ground in the region covered by the WNA surveys (Baumgartner et al. 2011, Prieto et al. 2014), whereas in the ENA, there is evidence that summer feeding grounds are located further north and west than the area surveyed (Sigurjónsson and Víkingsson 1997, Olsen et al. 2009). It is possible that sei whales sighted in our ENA study area were migrating individuals en route to feeding grounds, while most individuals in our WNA area were foraging. This could explain sei whale's negative relationship with PP in the ENA, but unimodal relationship in the WNA. Further telemetry studies at the basin scale are warranted to better elucidate cetacean habitat uses between regions.

2) Spatial population structure. One complicating consideration when exploring habitat relationships across an ocean basin is the organisms' spatial structuration into populations or stocks. Migratory baleen whales are known to exhibit significant stock structure across the North Atlantic Ocean (Anderwald et al. 2011, Prieto et al. 2012). Animals that belong to spatially distinct stocks may exhibit different habitat or prey specialization because they have learned to forage in separate environments (this may lead to local adaptation) (Louis et al. 2014). Such spatial stock structure could be responsible for diverging habitat relationships of cetaceans that belong to distinct stocks.

3) Flexibility in prey preferences. As mostly opportunistic predators, cetaceans may exhibit variable prey preferences that may express as spatial differences in modeled habitat relationships. Within the North Atlantic, fin whales mostly eat zooplankton in the Bay of Biscay (Spitz et al. 2018), but are well-known fish eaters on the northeast U.S. continental shelf (Kenney et al. 1997). Zooplankton and fish are expected to show distinct habitat relationships owing to their distinct trophic positions and physiology, in turn leading to distinct habitat relationships of predators that feed on them.

4) Changing correlations of proximal variables with 'true' variables. True factors driving cetacean distributions, such as forage prey availability, are intractable to measure over their vast distributional ranges. Consequently, variables widely available from remote sensing and assumed to be correlated with, or proxies for, these 'true' variables are traditionally used in SDMs (alternatively, prey variables derived from ecosystem model outputs such as SEAPODYM can be used (Roberts et al. 2016)). Correlations between these proximal and 'true' variables may change geographically (Dormann et al. 2012) because physical and biological processes enhancing prey availability may differ between regions (Benson et al. 2011), thereby leading to geographically inconsistent relationships.

5) Dissimilar available habitat between regions. Differing geomorphology, oceanic currents and environmental gradients between the two sides of a basin are expected to affect the spatial distribution and configuration of habitat available to cetaceans. In the WNA, the Gulf Stream – a western boundary current that brings warm water from the Gulf of Mexico and generates strong temperature gradients off the U.S. east coast – and the North Atlantic subtropical gyre – which creates a strong eastward decreasing productivity gradient (Tomczak and Godfrey 2003) – have no counterparts in the ENA. Thus, comparatively warmer and less productive waters not present in the ENA are available to cetaceans in the WNA and habitat relationships were estimated on wider ranges of SST and PP in the WNA. As habitat preferences are strongly tied to habitat availability (Beyer et al. 2010), dissimilar habitat availability between regions will likely affect modeled habitat relationships of cetaceans.

6) Other factors than habitat per se. For some species, differences in mean presence probabilities between regions appeared better captured with region-specific intercepts than with regions-specific relationship shapes (as revealed by the better fit of the RI model over the RS model; Table 1). These findings suggest that other factors than habitat preferences per se, such as anthropogenic pressures on cetacean populations, could be responsible for differences in mean presence probabilities between regions. Whales were widely hunted commercially until an international moratorium was declared in 1986, resulting in severe population depletions followed by a slow and geographically inconsistent recovery (Roman and Palumbi 2003). Present major threats to cetacean populations include accidental mortality in fishing gear, pollution and ship strikes (Avila et al. 2018). Past or present anthropogenic pressures heterogeneously distributed across a basin are expected to affect cetaceans differently over their distributional ranges. However, disentangling anthropogenic pressures from environment-driven effects on species distribution remains difficult, especially when histories of human pressures are unknown, or there is little information on the magnitude of their impacts (Yates et al. 2018).

A number of caveats applied to our analysis. First, for surveys conducted as far apart as the two sides of an ocean basin, differences in data collection are inevitable, owing to regional contexts, experience of survey organizations and availability of survey platforms. Although all included surveys followed a standard distance sampling methodology (Buckland et al. 2001), some differences in protocol occurred. For example, shipboard surveys in the WNA used high-powered binoculars, while shipboard surveys in the ENA used naked eyes. We corrected for these differences on the presence probability

of all cetacean taxa by including the effective area surveyed per segment (a measure of detectability) as an offset in our models. We caution that potential small errors in these offsets could lead to over- or under-estimation of presence probability on one side of the basin, with direct consequences on the magnitude of habitat relationships. As combining disparate survey datasets will remain a challenge (Virgili et al. 2019), we call for increased standardization of data collection protocols between both sides of the North Atlantic basin.

In addition, cetacean habitat relationships were estimated from summer data only and from a limited number of years. Further replications of surveys on both sides of the North Atlantic would be needed to assess how relationships vary across seasons and years. Incorporating surveys from larger latitudinal extents should also allow sampling broader ranges of environmental conditions experienced by cetaceans across their distributional ranges. Continuing data sharing (e.g. via the OBIS-SEAMAP repository (Halpin et al. 2006)) will be essential for continuing to collate high quality survey data on cetacean distributions across ocean basins.

Habitat relationships may be sensitive to sampling design in the species distribution data. In at-sea surveys of cetaceans and other marine species, poor weather conditions are one of the biggest factors that can prevent the realization of planned surveys, often for entire days or longer periods of time, leading to potentially unsystematical survey designs. In Supplementary material Appendix 7, we assessed the sensitivity of region-specific relationships to the survey design by applying jackknife procedures that excluded survey samples of 1 d or 7 d, refitting the models and examining the resulting relationships. In both regions, the shapes of resulting relationships were consistent with the shapes of overall relationships estimated from the full dataset, with very few of them falling outside the 95% confidence intervals of overall relationships (Supplementary material Appendix 7 Fig. A7-1, A7-2). These results underlined the robustness of habitat relationships to sampling design based on subsampling of both 1-d and 7-d survey periods.

Finally, habitat preferences of cetaceans result from complex mechanisms that are both environmentally-mediated (in response to resource availability) and behaviorally-mediated (arising from social, foraging or reproductive activities) (Palacios et al. 2013). These mechanisms cannot be fully elucidated using correlative SDMs, especially when survey datasets cover limited geographical extents. While SST and PP are classical dynamic predictors in cetacean SDMs (Roberts et al. 2016, Tobeña et al. 2016, Lambert et al. 2017, Mannocci et al. 2017, Pennino et al. 2017, Rogan et al. 2017, Fiedler et al. 2018, Becker et al. 2019, Chavez-Rosales et al. 2019, Virgili et al. 2019), they represent rather indirect predictors for these top predators, particularly for beaked whales and pilot whales that forage at depth below the photic zone (as noted by the estimated weak statistical effects in Fig. 2). Progress from correlative to process-based modeling approaches informed by directed in situ studies will ultimately uncover the mechanisms influencing cetacean habitat preferences (Palacios et al. 2013).

Conclusions

As large marine survey programs implemented over recent decades have opened up new possibilities for developing SDMs at basin scales, there is a risk for combining datasets from regions where species exhibit contrasting habitat relationships. Our statistical framework based on hierarchical GAMs allows testing for geographical variations in habitat relationships. Our findings revealed substantial differences in the relationships of cetaceans with three commonly included covariates across an ocean basin. Our proposed hypotheses provide a roadmap for better understanding geographical variations in the habitat relationships of cetaceans and other highly mobile marine species at the scale of ocean basins.

Data availability statement

The cetacean survey datasets used in this analysis are available from the OBIS-SEAMAP repository (<http://seamap.env.duke.edu/>).

References

Anderwald, P. et al. 2011. Possible cryptic stock structure for minke whales in the North Atlantic: implications for conservation and management. – Biol. Conserv. 144: 2479–2489.

Avila, I. C. et al. 2018. Current global risks to marine mammals: taking stock of the threats. – Biol. Conserv. 221: 44–58.

Baumgartner, M. F. et al. 2011. Diel vertical migration behavior of *Calanus finmarchicus* and its influence on right and sei whale occurrence. – Mar. Ecol. Prog. Ser. 423: 167–184.

Becker, E. A. et al. 2012. Forecasting cetacean abundance patterns to enhance management decisions. – Endanger. Species Res. 16: 97–112.

Becker, E. A. et al. 2019. Predicting cetacean abundance and distribution in a changing climate. – Divers. Distrib. 25: 626–643.

Behrenfeld, M. J. and Falkowski, P. G. 1997. Photosynthetic rates derived from satellite-based chlorophyll concentration. – Limnol. Oceanogr. 42: 1–20.

Benson, S. R. et al. 2011. Large-scale movements and high-use areas of western Pacific leatherback turtles, *Dermochelys coriacea*. – Ecosphere 2: 1–27.

Beyer, H. L. et al. 2010. The interpretation of habitat preference metrics under use–availability designs. – Phil. Trans. R. Soc. B 365: 2245–2254.

Buckland, S. T. et al. 2001. Introduction to distance sampling: estimating abundance of biological populations. – Oxford Univ. Press.

Burnham, K. P. and Anderson, D. R. 2002. Model selection and multimodel inference: a practical information–theoretic approach. – Springer Science and Business Media.

Byrne, M. E. et al. 2019. Behavioral response of a mobile marine predator to environmental variables differs across ecoregions. – Ecography 42: 1569–1578.

Cañadas, A. et al. 2002. Cetacean distribution related with depth and slope in the Mediterranean waters off southern Spain. – Deep-Sea Res. I 49: 2053–2073.

Cañadas, A. et al. 2018. The challenge of habitat modelling for threatened low density species using heterogeneous data: the case of Cuvier's beaked whales in the Mediterranean. – Ecol. Indic. 85: 128–136.

Chavez-Rosales, S. et al. 2019. Environmental predictors of habitat suitability and occurrence of cetaceans in the western North Atlantic Ocean. – Sci. Rep. 9: 1–11.

Croll, D. A. et al. 2005. From wind to whales: trophic links in a coastal upwelling system. – Mar. Ecol. Prog. Ser. 289: 30.

Dolomatov, S. et al. 2013. Role of temperature in regulation of the life cycle of temperate fish. – Russ. J. Mar. Biol. 39: 81–91.

Dormann, C. F. et al. 2012. Correlation and process in species distribution models: bridging a dichotomy. – J. Biogeogr. 39: 2119–2131.

Elith, J. and Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. – Annu. Rev. Ecol. Evol. Syst. 40: 677–697.

Fiedler, P. C. et al. 2018. Prediction of large whale distributions: a comparison of presence–absence and presence-only modeling techniques. – Front. Mar. Sci. 5: 1–15.

Gelman, A. and Hill, J. 2006. Data analysis using regression and multilevel/hierarchical models. – Cambridge Univ. Press.

Halpin, P. N. et al. 2006. OBIS-SEAMAP: developing a biogeographic research data commons for the ecological studies of marine mammals, seabirds and sea turtles. – Mar. Ecol. Prog. Ser. 316: 239–246.

Heerah, K. et al. 2017. Contrasting behavior between two populations of an ice-obligate predator in east Antarctica. – Ecol. Evol. 7: 606–618.

Kaschner, K. et al. 2006. Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. – Mar. Ecol. Prog. Ser. 316: 285–310.

Kaschner, K. et al. 2012. Global coverage of cetacean line-transect surveys: status quo, data gaps and future challenges. – PLoS One 7: e44075.

Kenney, R. et al. 1997. Estimates of prey consumption and trophic impacts of Cetaceans in the USA Northeast Continental Shelf ecosystem. – J. Northw. Atl. Fish. Sci. 22: 155–171.

Kery, M. and Royle, J. 2015. Applied hierarchical modeling in ecology: analysis of distribution, abundance and species richness in R and BUGS – volume 1. – Academic Press.

Laake, J. L. et al. 2014. Mark–recapture distance sampling. – Package mrds ver. 2.1.10.

Lambert, C. et al. 2017. How does ocean seasonality drive habitat preferences of highly mobile top predators? Part II: the eastern North-Atlantic. – Deep Sea Res. II 141: 133–154.

Louis, M. et al. 2014. Ecological opportunities and specializations shaped genetic divergence in a highly mobile marine top predator. – Proc. R. Soc. B 281: 20141558.

Mannocci, L. et al. 2015. Extrapolating cetacean densities beyond surveyed regions: habitat-based predictions in the circumtropical belt. – J. Biogeogr. 42: 1267–1280.

Mannocci, L. et al. 2017. Extrapolating cetacean densities to quantitatively assess human impacts on populations in the high seas. – Conserv. Biol. 31: 601–614.

Miller, J. A. 2012. Species distribution models: spatial autocorrelation and non-stationarity. – Prog. Phys. Geogr. Earth Environ. 36: 681–692.

Olsen, E. et al. 2009. First satellite-tracked long-distance movement of a sei whale *Balaenoptera borealis* in the North Atlantic. – Aquat. Mamm. 35: 313–318.

Palacios, D. M. et al. 2013. Beyond correlation: integrating environmentally and behaviourally mediated processes in models of marine mammal distributions. – Endanger. Species Res. 22: 191–203.

Pedersen, E. J. et al. 2019. Hierarchical generalized additive models in ecology: an introduction with mgcv. – PeerJ 7: e6876.

Pennino, M. G. et al. 2017. Habitat modeling for cetacean management: spatial distribution in the southern Pelagos Sanctuary (Mediterranean Sea). – Deep Sea Res. II 141: 203–211.

Prieto, R. et al. 2012. The forgotten whale: a bibliometric analysis and literature review of the North Atlantic sei whale *Balaenoptera borealis*: North Atlantic sei whale review. – Mammal Rev. 42: 235–272.

Prieto, R. et al. 2014. Sei whale movements and behaviour in the North Atlantic inferred from satellite telemetry. – Endanger. Species Res. 26: 103–113.

Redfern, J. V. et al. 2006. Techniques for cetacean–habitat modeling. – Mar. Ecol. Prog. Ser. 310: 271–295.

Redfern, J. V. et al. 2017. Predicting cetacean distributions in data-poor marine ecosystems. – Divers. Distrib. 23: 394–408.

Roberts, J. J. et al. 2010. Marine geospatial ecology tools: an integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB and C++. – Environ. Model. Softw. 25: 1197–1207.

Roberts, J. J. et al. 2016. Habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico. – Sci. Rep. 6: 22615.

Rogan, E. et al. 2017. Distribution, abundance and habitat use of deep diving cetaceans in the north-east Atlantic. – Deep Sea Res. II 141: 8–19.

Roman, J. and Palumbi, S. R. 2003. Whales before whaling in the North Atlantic. – Science 301: 508–510.

Schaub, M. et al. 2011. Habitat-density associations are not geographically transferable in Swiss farmland birds. – Ecography 34: 693–704.

Sequeira, A. M. M. et al. 2018. Transferring biodiversity models for conservation: opportunities and challenges. – Methods Ecol. Evol. 9: 1250–1264.

Sigurjónsson, J. and Víkingsson, G. 1997. Seasonal abundance of and estimated food consumption by Cetaceans in Icelandic and adjacent waters. – J. Northw. Atl. Fish. Sci. 22: 271–287.

Spitz, J. et al. 2018. Prey consumption by cetaceans reveals the importance of energy-rich food webs in the Bay of Biscay. – Prog. Oceanogr. 166: 148–158.

Tobeña, M. et al. 2016. Modeling the potential distribution and richness of Cetaceans in the Azores from fisheries observer program data. – Front. Mar. Sci. 3: 202.

Tomczak, M. and Godfrey, J. S. 2003. Regional oceanography: an introduction, 2nd ed. – Daya Publishing House.

Torres, L. G. et al. 2015. Poor transferability of species distribution models for a pelagic predator, the grey petrel, indicates contrasting habitat preferences across ocean basins. – PLoS One 10: e0120014.

Virgili, A. et al. 2019. Combining multiple visual surveys to model the habitat of deep-diving cetaceans at the basin scale. – Global Ecol. Biogeogr. 28: 300–314.

Whitehead, H. et al. 2008. Diversity of deep-water cetaceans in relation to temperature: implications for ocean warming. – Ecol. Lett. 11: 1198–1207.

Whitehead, H. et al. 2010. Diversity of deep-water cetaceans and primary productivity. – Mar. Ecol. Prog. Ser. 408: 1–5.

Whittingham, M. J. et al. 2007. Should conservation strategies consider spatial generality? Farmland birds show regional not national patterns of habitat association. – Ecol. Lett. 10: 25–35.

Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. – J. R. Stat. Soc. B 73: 3–36.

Wood, S. N. 2017. Generalized additive models: an introduction with R. – CRC Press Book.

Wood, S. N. 2018. Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation and GAMMs by REML/PQL. – Package mgcv ver. 1.8-26.

Wood, S. N. et al. 2016. Smoothing parameter and model selection for general smooth models. – J. Am. Stat. Assoc. 111: 1548–1563.

Worm, B. et al. 2005. Global patterns of predator diversity in the open oceans. – Science 309: 1365.

Yates, K. L. et al. 2018. Outstanding challenges in the transferability of ecological models. – Trends Ecol. Evol. 33: 790–802.

Yen, P. P. W. et al. 2004. Marine bird and cetacean associations with bathymetric habitats and shallow-water topographies: implications for trophic transfer and conservation. – J. Mar. Syst. 50: 79–99.

Zanini, F. et al. 2009. The transferability of distribution models across regions: an amphibian case study. – Divers. Distrib. 15: 469–480.

Zuur, A. F. et al. 2010. A protocol for data exploration to avoid common statistical problems. – Methods Ecol. Evol. 1: 3–14.

Supplementary material (available online as Appendix ecog-04979 at <www.ecography.org/appendix/ecog-04979>). Appendix 1–7.