Combining multiple visual surveys to model the habitat of deep-diving cetaceans at the basin scale

Large scale modelling of deep-diving cetacean habitats

Auriane Virgili [1]*, Matthieu Authier [2], Oliver Boisseau [3,4], Ana Cañadas [5], Diane Claridge [6], Tim Cole [7], Peter Corkeron [7], Ghislain Dorémus [2], Léa David [8], Nathalie Di-Méglio [8], Charlotte Dunn [6], Tim E. Dunn [9], Isabel García-Barón [10], Sophie Laran [2], Giancarlo Lauriano [11], Mark Lewis [9], Maite Louzao [10], Laura Mannocci [12,a], José Martínez-Cedeira [13], Debra Palka [7], Simone Panigada [14], Emeline Pettex [2,b], Jason J. Roberts [12], Leire Ruiz [15], Camilo Saavedra [16], M. Begoña Santos [16], Olivier Van Canneyt [2], José Antonio Vázquez Bonales [5], Pascal Monestiez [1,17], Vincent Ridoux [1,2]

[1] Centre d'Etudes Biologiques de Chizé - La Rochelle, UMR 7372 CNRS - Université de La Rochelle, Institut du Littoral et de l'Environnement, 17000 La Rochelle, France; [2] Observatoire PELAGIS, UMS 3462 CNRS - Université de La Rochelle, Systèmes d'Observation pour la Conservation des Mammifères et des Oiseaux Marins, 17000 La Rochelle, France; [3] Marine Conservation Research, 94 High Street, Kelvedon, CO5 9AA, UK; [4] Song of the Whale research team, International Fund for Animal Welfare (IFAW), 87-90 Albert Embankment, London, SE1 7UD, UK; [5] Alnilam Research and Conservation, Pradillos 29, 28491-Navacerrada, Madrid. Spain; [6] Bahamas Marine Mammal Research Organisation, P.O. Box AB-20714, Marsh Harbour, Abaco, Bahamas; [7] Protected Species Branch, NOAA Fisheries Northeast Fisheries Science, 166 Water St Woods Hole Massachusetts, 02543 USA; [8] EcoOcéan Institut, 34090 Montpellier, France; [9] Joint Nature Conservation Committee, Inverdee House, Baxter Street, Aberdeen, AB11 9QA, UK; [10] AZTI, Herrera Kaia, Portualdea z/g, 20110 Pasaia, Spain; [11] Institute for Environmental Protection and Research – ISPRA, via V. Brancati 60, 00144 Roma, Italy; [12] Marine Geospatial Ecology Laboratory, Duke University, Durham, North Carolina, USA; [13] CEMMA, Camiño do Ceán, n°2, 36350, Nigrán, Pontevedra, Spain; [14] Tethys Research Institute, Acquario Civico, 20121 Milano, Italy; [15] AMBAR Elkartea organisation, Mungia Bidea 9, 3B, 48620 Plentzia, Bizkaia, Spain; [16] Instituto Español de Oceanografía, Centro Oceanográfico de Vigo, 36390 Vigo, Spain; [17] BioSP, INRA, 84914 Avignon, France.

Corresponding author: Auriane Virgili, 2 rue Olympe de Gouges 17000 La Rochelle, France
auriane.virgili@univ-lr.fr

[a] *Current address:* UMR MARBEC (IRD, Ifremer, Univ. Montpellier, CNRS), Institut Français de Recherche pour l'Exploitation de la Mer, Av. Jean Monnet, CS 30171, 34203 Sète, France

[b] *Current address:* Cohabys - ADERA, Université de La Rochelle 2, rue Olympe de Gouges, Bâtiment ILE, 17000 La Rochelle

Acknowledgments

We are grateful to the many observers who participated in the surveys and collected all the data and also to the ships' captains, crews and pilots. We thank Phil Hammond and his team for providing SCANS and CODA survey data. THUNNUS survey was carried out thanks to the collaboration of the General Directorate of Fisheries and Maritime Affairs, Government of Galicia. We thank the *Direction Générale de l'Armement* (DGA), including Odile Gérard and Carole Nahum, for funding Auriane Virgili's doctoral research grant. ML was funded by a Ramón y Cajal (RYC-2012-09897) postdoctoral contract of the Spanish Ministry of Economy, Industry and Competitiveness, whereas IGB was supported by a PhD fellowship (FPI, BES-2014-070597) of the Spanish Ministry of Economy, Industry and Competitiveness. This study is a contribution to the CHALLENGES (CTM2013-47032-R) project of the Spanish Ministry of Economy, Industry and Competitiveness. EcoOcéan Institut acknowledge its partners for the participation in the collection of data at sea: École Pratique des Hautes Études, WWF-France, Swiss Cetacean Society, Cybelle Planète, Participe Futur and Fondation Nicolas Hulot. This work is linked to the program "PELAGIC 2" lead by the CESAB which aims to establish globally relevant areas for the conservation of marine mammals. We finally thank Charlotte Dunn and Tim Dunn for having proofread the English of the manuscript and the anonymous reviewers for their very helpful comments that led to a clearer and much improved manuscript.

Biosketch

This work is part of Auriane Virgili's PhD project which aims to model distributions of rare marine species with a focus on deep-diving cetaceans. These species are rare and difficult to observe at the surface thus it was necessary to assemble datasets from different surveys to model their distribution in the North Atlantic Ocean and the Mediterranean Sea. This required the collaboration of many organisations represented by the different co-authors of this article.

DR. AURIANE VIRGILI (Orcid ID : 0000-0002-7581-7985)

Article type : Research Papers

Combining multiple visual surveys to model the habitat of deep-diving cetaceans at the basin scale

Large scale modelling of deep-diving cetacean habitats

Abstract

Aim: Deep-diving cetaceans are oceanic species exposed to multiple anthropogenic pressures including high intensity underwater noise, and knowledge of their distribution is crucial to manage their conservation. Due to intrinsic low densities, wide distribution ranges and limited presence at the sea surface, these species are rarely sighted. Pooling data from multiple visual surveys sharing a common line-transect methodology can increase sightings but requires accounting for heterogeneity in protocols and platforms.

Location: North Atlantic Ocean and Mediterranean Sea

Time period: 1998 to 2015

Major taxa: *Ziphiidae*; *Physeteriidae*; *Kogiidae*

Methods: About 1,240,000 km of pooled effort provided 630 sightings of ziphiids, 836 of physeteriids and 106 of kogiids. For each taxon, we built a hierarchical model to estimate the effective strip width depending on observation conditions and survey types. We then modelled relative densities in a Generalised Additive Modelling framework. Geographical predictions were limited to interpolations identified with a gap analysis of environmental space coverage.

Results: Deeper areas of the North Atlantic gyre were mostly environmental extrapolation, thereby highlighting gaps in sampling across the different surveys. For the three species groups, the highest relative densities were predicted along continental slopes, particularly in

the western North Atlantic Ocean where the Gulf Stream creates dynamic frontal zones and eddies.

Main conclusions: Pooling a large number of surveys provided the first basin-wide models of distribution for deep-diving cetaceans, including several data-deficient taxa, across the North Atlantic Ocean and the Mediterranean Sea. These models can help the conservation of elusive and poorly known marine megafauna.

Keywords: Beaked whales, Data-assembling, Deep-diving cetaceans, Habitat modelling, Kogiids, Sperm whales

1. Introduction

Deep-diving cetaceans, defined here as beaked whales (family *Ziphiidae*; *e.g. Ziphius cavirostris*, *Hyperoodon* spp. and *Mesoplodon* spp.) and sperm whales (families *Physeteridae* and *Kogiidae*), are distributed worldwide. They are oceanic species that feed in deep waters during long dives (close to or even longer than an hour; Perrin et al., 2009). Due to their offshore habitat and the short time they remain available at the sea surface, little is known about their synoptic distribution (especially for kogiids and ziphiids). Moreover, these species are threatened by anthropogenic activities, including bycatch, debris ingestion, ship collisions (Carrillo & Ritter, 2010; Madsen et al., 2014; Unger et al., 2016) and any activity producing high intensity noise (*e.g.* military sonars, seismic guns or techniques used on large maritime construction projects; Stone & Tasker, 2006). Recent studies have demonstrated the sensitivity of deep-diving cetaceans, and particularly beaked whales, to underwater noise pollution, with a number of unusual stranding events associated with the use of military sonars (Fernández et al., 2005; D'Amico et al., 2009). To mitigate the impact of these activities, accurate knowledge of the distribution and abundance of deep-diving cetaceans is crucial to Marine Spatial Planning to inform management measures at a national scale (Douvere, 2008). International initiatives, such as Important Marine Mammal Areas (IMMAs, Corrigan et al., 2014), are needed for these highly mobile species. However, any single survey often yields only a handful of sightings that are then restricted to areas too small compared to the large geographical scale needed for effective conservation planning.

Data-assembling is increasingly used to model habitat preferences of cetaceans at the basin scale (Roberts et al., 2016; Rogan et al., 2017; Cañadas et al., 2018). Due to the various protocols, platform types and observation heights, species detectability and data quality vary with surveys. In addition, each survey may not collect the same information, particularly with regard to observation conditions. Some surveys only record Beaufort sea-state while others record additional parameters that also influence species detection, such as sun glare, cloud coverage or wave height. In the process of synthesising different datasets, only variables common across all datasets can generally be retained in a broad scale analysis, which nevertheless needs to account for heterogeneity. Finally, to make basin-wide predictions from the assemblage of a number of local surveys, identifying areas of environmental extrapolations is crucial to bolster confidence in predicted maps (Mannocci et al., 2018).

Our study aims to understand how deep-diving cetaceans are distributed at a large scale and to highlight areas of high relative densities for conservation purposes. To model the habitats of deep-diving cetaceans at a large scale, we assembled data from different surveys in the North Atlantic Ocean and the Mediterranean Sea from 15 organisations. To take into account heterogeneity in sighting protocols, we built a hierarchical model to estimate the effective strip width across platforms and observation conditions. We then modelled relative densities of three deep-diving cetacean taxa with Generalized Additive Models (GAM). Finally, we performed a gap analysis (Jennings, 2000, Mannocci et al., 2018) to assess the reliability of the predictions outside the surveyed area.

2. Methods

2.1. Data origin

The study area encompassed the North Atlantic Ocean and the Mediterranean Sea from the Guiana Plateau to Iceland, excluding the Baltic and Black Seas, the Gulf of Mexico and the Hudson Bay, both because of an absence of effort data and of ecological and environmental differences (Fig. 1A; Appendix S1 in Supporting Information). Four sub-regions were defined in the study area (Table 1; Fig. 1A): the northeast Atlantic Ocean (NE-ATL), the northwest Atlantic Ocean (NW-ATL), the tropics and the Mediterranean Sea (MED).

We assembled visual shipboard and aerial surveys performed by 15 independent organisations in the North Atlantic Ocean and the Mediterranean Sea between 1998 and 2015 (Fig. 1; survey-specific information are detailed in Appendix S2 in Supporting Information). Except for the JNCC-ESAS surveys that use a 300m-strip-transect methodology, all surveys used line-transect methodologies that correct for non-detection bias with the estimation of an Effective Strip Width (ESW) from the measurement of the perpendicular distances to the sightings (Buckland et al., 2015; see below).

To account for the difficulty in identifying deep-diving cetaceans to the species level (*e.g.* genera *Mesoplodon*, *Kogia*), we pooled species into three groups: (1) beaked whales, consisting of Cuvier's beaked whales (*Ziphius cavirostris),* mesoplodonts (*Mesoplodon* spp.) and northern bottlenose whales *(Hyperoodon ampullatus*), (2) sperm whales (*Physeter* *macrocephalus*), and (3) kogiids, including pygmy (*Kogia breviceps)* and dwarf sperm whales *(K. sima*).

2.3. Data processing

2.3.1. Data-assembling

All survey datasets were standardised for units and formats (*e.g.* date, time and coordinates) and aggregated into a single common dataset. A specific coordinate projection encompassing the entire survey area was used for accurate distance computations (Albers equal-area conic defined from http://projectionwizard.org). Effort data were linearized and divided into 5 km segments using ArcGIS 10.3 (ESRI, 2016) and the Marine Geospatial Ecology Tools software (Roberts et al., 2010). The segment length represented a trade-off value across varying survey transect lengths, for example aerial surveys had transect lengths of up to 100 km while shipboard surveys were often much shorter. Finally, for each species group, sightings were linked to their respective 5 km segments.

Encounter rates were calculated in each sub-region as: $(number\ of\ encounters/total\ distance\ travelled) * 100$.

2.3.2. Environmental variables

In habitat models, we tested the static and dynamic variables that were expected to influence the distributions of deep divers (Table 2). All variables were resampled at a 0.25° resolution because of the very large size of the study area and the spatial resolution of the variables (Table 2; Appendix S3 in Supporting Information). Spatial gradients of sea surface temperature (SST) were calculated as the difference between the minimum and maximum SST values in an eight-pixel buffer around a given pixel. Net primary production (NPP) was used as a proxy for prey availability.

Dynamic variables, which relate to the movements of water masses or prey availability, were computed at a monthly resolution *i.e.* averaged over the 29 days prior to each sampled day to avoid gaps in remote sensing oceanographic variables. They were used in addition to static variables because they reveal the presence of time-stable structures such as temperature gradients or eddies when variables are averaged.

2.3.3. Effective Strip Width estimation

Line-transect surveys are commonly used to estimate cetacean abundance (Hammond et al., 2013; Buckland et al., 2015). A key parameter to estimate this abundance is the effective strip width (ESW) which corrects the decreasing detection of animals with distance from the trackline. ESW is expected to depend on survey platform height, platform type, sea-state, species, etc… (Buckland et al., 2015).

ESW estimation was a key step in the data-assembling process to take into account heterogeneity in effort per segment in the models and to directly compare the different surveys (Hedley & Buckland, 2004). ESWs are generally estimated for each survey (*i.e.* no pooling of information) by using the 'Distance' software (Thomas et al., 2010; Buckland et al., 2015). However, for deep-diving cetaceans, the majority of surveys contained insufficient sightings to allow survey-specific detection functions to be fitted. Consequently, for each species group, we pooled sightings from the various surveys, taking into account survey heterogeneity. We built a hierarchical model in which survey identity was included as a random effect.

In conventional distance sampling (Marques & Buckland, 2003; Buckland et al., 2015), factors such as the characteristics of the species being surveyed, search methods, search platform, environmental conditions can all affect ESW estimation. However, the different datasets did not always contain this information, especially regarding observation conditions. All surveys recorded environmental data such as Beaufort sea-state, cloud coverage and sun glare, although Beaufort sea-state was the only parameter recorded by all of them. Platform type, observation height and Beaufort sea-state were used as covariates in the hierarchical model.

Truncation distance w was first determined as the 95th percentile of the set of perpendicular distances for each species group, *i.e.* the 5% most distant sightings were discarded from the analysis (Buckland et al., 2001, page 16). Then, we created classes to pool the different surveys; namely platform type (plane or boat), observation height (*e.g.* 0-5 m; 5-10 m…) and Beaufort sea-state (0-1; 1-2; 2-3 and 3-4; data collected beyond a Beaufort sea-state 4 being removed from the analysis). Hierarchical modelling was then performed in R-3.3.1 (R Core Team, 2016) in a Bayesian framework using JAGS version 4-6 and package 'rjags' (jags model in Appendix S4 in Supporting Information; Royle & Dorazio, 2008; Plummer, 2016).

For each taxa, perpendicular distances of sightings were used to estimate a detection function with a hazard key. For a sighting i made during survey s at height j under Beaufort sea-state k, let d_{jks}^{i} denotes the perpendicular distance. The detection probability of sighting i is:

$$\begin{cases} p_{ijk}^{s} = g_s(d_{ijk}) = 1 - \exp\left(-\left(\frac{d_{ijk}}{\sigma_{jks}}\right)^{-\nu_s}\right) \\ \log(\sigma_{jks}) = \beta_{j0} + \beta_{j1} \times k + \alpha_s \end{cases}$$

where β_{j0} and β_{j1} are respectively random intercept and slope parameters for the effect of platform height; and α_s and ν_s are survey random effects. Bivariate random effects were specified with a Cholesky decomposition and using priors for the Cholesky factors from Kinney & Dunson (2008). We used half Student-t distributions with 3 degrees of freedom and scale set to 1.5 as priors for dispersion parameters, and standard normal priors for all other parameters. Four chains were run with a warmup of 10,000 iterations, followed by another 10,000 iterations (with a thinning factor of 10). Parameter convergence was assessed with Gelman-Rubin $\hat{R}$ statistics. Posterior inferences were based on the pooled sample of 4,000 values (1,000 per chain).

The advantage of setting a hierarchical model to estimate detection functions is to borrow strength across the different datasets to increase the precision of estimates. For each combination of survey – platform type – observation height – Beaufort sea-state, estimated detection functions are shrunk towards a common detection function (itself estimated from the data) according to the available data corresponding to this particular combination of survey – platform type – observation height – Beaufort sea-state. If, for a given combination of parameters, there were few sightings, the estimated detection function was very close to the common detection function, whereas if there were enough data, the estimated detection function could deviate from this common function. Upon model fitting and successful parameter estimation, the ESW for each combination of survey – platform type – observation height – Beaufort sea-state was computed:

$$\mathrm{ESW}_{jks} = \int_0^w g_s(x)dx = \int_0^w \left[1 - \exp\left(-\left(\frac{x}{e^{\beta_{j0}+\,\beta_{j1}\times k+\,\alpha_s}}\right)^{-\nu_s}\right)\right]dx$$

The posterior mean value of estimated ESW was then allocated to each segment with respect to species group, survey, platform type, sea-state and observation height class.

2.4. Habitat modelling

To model habitat preferences of deep-divers, we fitted Generalised Additive Models (GAMs; Hastie & Tibshirani, 1986; Wood, 2006) with a Tweedie distribution to account for over-dispersion (Foster & Bravington, 2013) with the 'mgcv' R-package (R-3.3.1. version; Wood, 2013). GAMs extend Generalised Linear Models to allow for smooth nonlinear functions of predictor variables (Hastie & Tibshirani, 1986; Wood, 2006). The mean number of individuals per segment μ was modelled with a logarithmic link function:

$$\log(\mu) = \alpha + \sum_p f(X_p)$$

where $f(X_p)$ are non-parametric smooth functions (thin plate regression splines) of the covariates and α is the intercept (Hastie & Tibshirani, 1986). To attenuate the scope for over-fitting, the maximum number of knots was limited to 4 (mgcv parameter k = 4; Wood, 2006). An offset equal to segment length multiplied by twice the ESW was included (except for the JNCC-ESAS surveys in which only one side of the vessel was surveyed). We removed combinations of variables with Spearman partial correlation coefficients higher than |0.7| (Dormann et al., 2013; Mannocci et al., 2014) and tested all models with combinations of one to four variables. A maximum of four covariates per model was used to avoid excessive complexity of models and difficulty in their interpretation (Mannocci et al., 2014). Model selection was done with the Akaike Information Criterion (AIC, the lower the better; Anderson & Burnham, 2002) and Akaike model weight (akaike.weights function from 'qpcR' package; Spiess, 2014).

A key assumption of line-transect surveys is that animals on the trackline are always detected (Buckland et al., 2001). However, this assumption is not met with diving species and trackline detection probability $g(0)$ needs to be accounted for (Barlow, 2015). Observers on a plane spend less time in a given area and the following inequality is expected: $g^{\text{boat}}(0) > g^{\text{plane}}(0)$. Thus a segment of effort with zero sighting of deep-divers is more likely to be a false absence (non-detection of a diving animal present on the trackline) if that segment comes from a plane survey rather than a boat survey. As detection probability $g(0)$ was not available for every survey and is expected to differ between platforms, we calculated the ratio of $g(0)$ between the plane and boat platforms from Roberts et al. (2016) and obtained a ratio of approx. 1/5 for beaked whales, approx. 2/5 for sperm whales and approx. 1/3 for kogiids. These crude ratios were then used to weight plane segments with zero sightings when fitting GAMs. While this method does not fully correct for availability bias, it down-weights zeroes from plane surveys.

We fitted "year-round" models as the studied taxa have been reported to show little or no seasonal variation in their habitats (*e.g.* Wimmer & Whitehead, 2004; McSweeney et al., 2007). We did not model yearly variations because of little temporal overlap between surveys. Consequently, the year effect is confounded with survey heterogeneity.

Predictions of relative densities (in number of animals per km^2) were provided at 0.25° resolution. There was not enough data to fit a model by month or by season (the number of sightings in winter was too low) and we therefore produced averaged maps over the entire time period. These predictive maps provided the expected distribution of beaked whales, sperm whales and kogiids according to static and monthly environmental conditions to highlight relationships with static (canyons and seamounts) and time-stable structures (temperature gradients or eddies).

Finally, coefficients of variation (CVs) were estimated for each 0.25° pixel. Coefficients of variations are a measure of the prediction uncertainty per cell, it is a standard error associated with the calculation of the prediction. Therefore, high CVs indicate high model uncertainties due to the lack of detection.

2.5. Gap analysis

Even though more than 1,240,000 km of effort was pooled, extensive geographical gaps remained. Predictions in the middle of the Atlantic Ocean are from geographical extrapolation (Fig. 1A) but not necessarily environmental extrapolations. The latter depends on the selected habitat models and covariates therein. We conducted a gap analysis on environmental space coverage to identify areas where habitat models could produce reliable predictions outside survey blocks, *i.e.* geographical extrapolation, whilst remaining within the ranges of surveyed conditions for the combinations of covariates selected by the models, *i.e.* areas of environmental interpolation (Jennings, 2000; Mannocci et al., 2018).

From the selected models for each taxa, we estimated the convex hull defined by the environmental data used to fit habitat models (hereafter the calibration data). The convex hull of a set of points is the smallest convex envelop that contains all these points. We then assessed whether a prediction from a set of environmental covariates with a given model fall inside or outside this convex hull (King & Zeng, 2007; Authier et al., 2016). We used climatological predictors instead of monthly predictors to lessen the computational burden.

Due to the large number of data (more than 280,000 points in the calibration dataset), convex hulls were estimated by random sub-sampling with the 'WhatIf' R-package (Stoll et al., 2014). We randomly extracted a fraction of the calibration dataset (10,000 points) to estimate a convex hull and assess environmental extrapolation in the prediction dataset. A combination of climatological predictor values that fall inside the convex hull corresponds to an interpolation. Combinations of climatological predictor values that were classified as interpolations were set aside but other combinations were retained and further tested against another random sample of 10,000 points from the calibration data. This procedure was carried out until the complete calibration dataset was examined.

The full procedure was conducted twice. In a simple approach, the full range of sampled variables was considered to identify all points of the whole study area where the actual combinations of environmental variables had been sampled in survey blocks. In a more 'precautionary approach', we excluded 5% of the extreme values of the sampled environmental variables to include in the interpolation areas only the points whose associated combinations of covariates fell within 95% of the core ranges sampled. This allowed the definition of two levels of confidence (hereafter 'simple' and 'precautionary') in the predictions.

Finally, we produced maps delineating the extent of the simple and precautionary interpolation areas and overlaid them with the relative density prediction maps to show areas with greater reliability.

3. Results

3.1. Encounter rates

The survey pool represented a total of 1,240,000 km of on-effort transects (*i.e.* following a transect at a specified speed and altitude with a specified level of visual effort) of which 58% were carried out by plane and 42% by boat (Fig. 1A, Table 1). Effort data with a Beaufort sea-state higher than 4, which represented 9% of the effort data, were removed from further analysis to only keep sightings collected during good to excellent detection conditions. Most sampling effort was performed in the northeast (37 %) and northwest (45 %) Atlantic Ocean. Surveys in the Mediterranean Sea and in the tropics represented respectively only 16 % and 2 % of total sampling effort.

A total of 630 sightings of beaked whales, 836 sightings of sperm whales and 106 sightings of kogiids, mainly distributed in the northeast and northwest Atlantic Ocean (north of the 35°N latitude) and in the northwest Mediterranean Sea, were assembled for the present study (Fig. 1B-D).

Overall encounter rates were very low with 0.05 sightings·100 km^{-1} for beaked whales, 0.07 sightings·100 km^{-1} for sperm whales and <0.01 sightings·100 km^{-1} for kogiids (Table 3). The highest encounter rates were recorded in the tropics for all three species groups, particularly for kogiids. There were no sightings of kogiids in the Mediterranean Sea.

3.2. Effective strip width

Estimated ESWs varied across surveys and platform type and were on average narrower in aerial than shipboard surveys (Fig. 2). This is probably because aerial observers are more restricted to recording animals below the plane while shipboard observers can look further afield. ESWs were generally larger and more consistent between surveys using the same platform type, for sperm whales than for beaked whales. There were not enough kogiid sightings to estimate an ESW for each survey and particularly for shipboard surveys; consequently, we pooled all aerial surveys and estimated an ESW of 1.1 km that was then applied to all surveys (shipboard and aerial).

The outcomes from the hierarchical model were consistent with expectations (Fig. S4.1 in Supporting Information S4): a decrease in Beaufort sea-state (less wind-sea) resulted in a larger ESW (milder non-detection bias).

3.3. Habitat modelling

For each species group, selected variables, explained deviances and Akaike weights are shown in Table 4.

Beaked whales

Highest relative densities were found in depth *ca.* 1,500 m, high values of slopes and SST and intermediate NPP. This resulted in high predicted relative densities of beaked whales along steep slope areas associated with deep depths and high gradients of temperature, particularly on the western side of the Atlantic Ocean. The lowest relative densities were predicted in the Mediterranean Sea (Fig. 3B).

The gap analysis identified areas where the combination of the four variables selected by the best model had not been sampled. Reliable predictions were available for 94% of the study area under the simple approach and only 53% under the precautionary approach (Fig. 3B and 3C). This discrepancy was mostly due to low sampling effort in the oceanic zone. Coefficients of (temporal) variation were higher on the continental shelf associated with high gradients of SST, where beaked whales were not sighted in any of the surveys (Fig. S5.2A in Supporting Information S5).

Sperm whales

Predicted relative densities of sperm whales increased in deep waters (> 2000 m) associated with high gradients of SST and high NPP. The highest relative densities were also predicted on the western side of the Atlantic Ocean, along the Gulf Stream, although were lowest in the Mediterranean Sea (Fig. 4B).

Reliable predictions for sperm whales were available for 84% of the study area under the simple approach and only 30% under the precautionary approach because of low survey effort in deeper areas. The highest predicted relative densities were predicted outside the precautionary interpolation zone (Fig. 4B and 4C). Coefficients of (temporal) variation were highest in non-sampled areas where uncertainty was therefore greatest (Fig. S5.2B in Supporting Information S5).

Kogiids

As the Akaike weight was small for kogiids (0.17), we used model-averaging and generated predictions from the five first models (cumulative Akaike weight of 0.63) and because all predictions were very similar (see Appendix S6 in Supporting Information), we only kept the first model for practical reasons. The highest relative densities were found in deep waters associated with fronts, canyons and seamounts (Fig. 5B). The highest relative densities were predicted on the western side of the Atlantic Ocean, along the Gulf Stream (Fig. 5C).

Reliable predictions for kogiids were available for 94% of the study area under the simple approach against only 55% under the precautionary approach because of low survey effort in deeper areas (Fig. 5C). Coefficients of (temporal) variation were highest in shallow waters and in the Mediterranean Sea where kogiids were not sighted in any of the surveys (Fig. S5.2C in Supporting Information S5).

4. Discussion

Deep-diving cetaceans are species characterised by low sighting rates and modelling their habitats is particularly challenging. Our study pooled different surveys allowing us to capitalise on more than 1,240,000 km of survey effort deployed over the North Atlantic Ocean and the Mediterranean Sea in the past two decades. For each taxon, we built a hierarchical model to estimate the effective strip width depending on observation conditions and surveys. We investigated habitats of deep-divers using GAMs with a focus on quantifying how reliable the predictions were. The selected habitat models of deep-diving cetaceans included static environmental variables such as depth and slope as well as spatial gradients of temperatures, revealing the highest densities in the western North Atlantic Ocean. Deeper areas of the North Atlantic gyre were mostly areas of environmental extrapolation, thereby highlighting gaps in sampling across the different surveys.

4.1. Methodological considerations

Over the past few years, data-assembling has been increasingly used for the study of top marine predators (Roberts et al., 2016; Rogan et al., 2017; Cañadas et al., 2018). Due to the very low sighting rates of deep-diving cetaceans, each survey taken separately cannot provide enough data to investigate the habitats of these rare species. In contrast to Rogan et al. (2017), we did not assemble data collected with similar protocols but data collected with different variants of the line transect distance sampling protocol which meant standardising the data according to their core communalities before developing a single spatial model. Ideally, at a time when shared databases are becoming increasingly important (*e.g.* OBIS SEAMAP -- http://seamap.env.duke.edu/, EMODnet -- http://www.emodnet.eu/), implementing standardised survey methods would greatly improve data compatibility, by enhancing the level of communalities in shared datasets, and helping to describe large-scale habitats and distributions of marine species. However, we realise this can lead to financial and logistical constraints and the work we present here could be a way to embrace and incorporate the diversity of data collection methods.

Hierarchical modelling accommodates heterogeneity between surveys; it borrows strength across surveys ('partial pooling') when estimating survey-specific ESWs. The resulting estimates are biased (in proportion to the available data contributed by each survey) toward a common mean, although are more precise than those that would be obtained if each survey was analysed separately ('no pooling' scenario) as it is usually done when the number of sightings per survey is large (Buckland et al., 2015; Laran et al., 2017; Redfern et al., 2017). Results from the hierarchical model were consistent with expectations and showed that a decrease in Beaufort sea-state values resulted in increased ESW estimates.

The majority of environmental variables we used in habitat modelling describe the euphotic zone (upper layer) because variables that describe the deep-water column are difficult to obtain or simply do not exist at a basin-wide scale. As deep-diving cetaceans spend most of their time at depth and generally feed on mesopelagic to bathypelagic prey (*e.g.* Perrin et al., 2009; Spitz et al., 2011), the use of surface variables limits the ability to correctly infer their habitat. The identified relationships between deep-diving cetacean abundance and environmental variables may be indirect rather than causal (Austin, 2006). Although causation may be out of reach, prediction remains a worthy goal, especially for spatial planning and conservation (McShea, 2014).

We took care in using appropriate statistical tools for modelling the habitat of species with few sightings (Virgili et al., 2018). Indeed, Virgili et al. (2018) showed that GAMs with a Tweedie distribution generated reliable habitat modelling predictions for rarely sighted marine predators. Here, the habitat models we selected had moderate to high levels of explained deviances (from 20.6% to 55.7%), suggestive of a good fit to the data. Nevertheless, the rather high explained deviance of the kogiid model (55.7%) might indicate some level of model over-fitting due to the small dataset, even if predictions were in general consistent with the known ecology of the species group (McAlpine, 2009).

4.2. Large-scale deep-diver habitats

Depth and spatial gradients of sea surface temperature were consistently selected across deep-diving cetaceans, suggesting a major influence of topographic features and thermal fronts in structuring their habitats. As a result, higher relative densities of deep-divers were predicted in areas of strong gradients associated with thermal fronts in which deep-diver prey aggregates (Bost et al., 2009; Woodson & Litvin, 2015). Indeed, deep-divers typically feed on mesopelagic to bathypelagic species, such as pelagic cephalopods and benthic fishes (Spitz et al., 2011) that aggregate along continental slopes where temperature gradients are the strongest. Hence, the Gulf Stream, which is the most active frontal zone in the study area compared to the eastern boundary currents that are broader and much slower, may explain the high predicted relative densities of deep-divers on the western side of the North Atlantic Ocean (Waring et al., 2001; Roberts et al., 2016).

Despite commonalities, each studied taxon also showed specificities. Slope appeared to be an important predictor of beaked whale relative density. The prey targeted by beaked whales are more specific than those of sperm whales, which have broader prey size spectrum (Spitz et al., 2011), and their distribution is more driven by dynamic variables than by static features. Accordingly, the selected model for sperm whales included more dynamic variables such as NPP and SSH than for beaked whales. Canyons and seamounts were included in the selected model for kogiids, suggesting a more restricted habitat than for the other two groups of deep-divers conforming Staudinger et al.'s (2014) evidence of how kogiids' feeding areas concentrated on the deeper shelf and slope, particularly in the epipelagic and mesopelagic zones.

Overall, our model predictions corroborated species distribution predictions of previous smaller-scale studies. In the Mediterranean Sea, our predictions were consistent with the documented presence of beaked whales and sperm whales in the Alborán, Tyrrhenian and Ligurian Seas (Praca & Gannier, 2008; Arcangeli et al., 2015; Lanfredi et al. 2016; Cañadas et al., 2018) and along the eastern coasts of the Mediterranean Sea (Podestà et al., 2006). In the North Atlantic Ocean, the highest relative densities of beaked whales and sperm whales were predicted along the slope, a result consistent with those of Rogan et al.'s (2017) and Roberts et al.'s (2016). In the Northwest Atlantic Ocean, higher relative densities of kogiids were predicted in warmer and deeper waters, which is consistent with their known ecology (McAlpine, 2009) and the predictions of Mannocci et al. (2017) except for predictions off the coast of Florida. Our predictions could probably be improved by incorporating the NOAA SEFSC surveys of southeast US waters off Florida and Virginia. In contrast to beaked and sperm whales, we were not able to fit a hierarchical model on kogiid sightings and resorted to complete pooling of the plane data to estimate an ESW. This shortcoming probably resulted in a larger bias (with respect to the true density) in predicted relative density of kogiids compared to other deep-diving species. Given the paucity of information on kogiids, we think that our results are tentative but important nonetheless.

The gap analysis revealed large gaps in environmental space coverage across the study area, especially in the deeper and less productive waters of the central north Atlantic gyre and in tropical waters. High relative densities of deep-divers were predicted at the margin of the precautionary interpolation zone (Figs. 3-5) in particular because deeper waters and steeper slopes were within the upper 2.5% quantiles of aggregated survey coverage for these two physiographic covariates. This suggests that sampling effort was not sufficient in deeper and steeper areas and more intensive sampling effort performed in these areas could help better describe the habitat used by deep-divers.

4.3. Management considerations

The management and conservation of species and ecosystems increasingly relies on habitat models (McShea, 2014; Hazen et al., 2016). The ability of these to predict species occurrence in non-sampled or poorly documented areas is useful (Fleishman et al., 2001; Lumaret & Jay-Robert, 2002) because the implementation of dedicated surveys is sometimes impracticable due to budgetary and logistical challenges. It is logistically challenging to carry out dedicated cetacean surveys in the middle of the North Atlantic Ocean. However, by collecting data on both sides of the Atlantic Ocean, relative density maps were produced and our analyses indicated these predictions may be reliable (Figs. 3D, 4D, 5D).

Here, we showed that deep-diving cetaceans are closely associated with stable topographic features, thus it could be possible to delineate marine protected areas that cover the principal habitats used by the species (*e.g.* Cañadas et al., 2005). However, these species are also responsive to temporally dynamic structures, such as thermal fronts, implying that protected areas will need to be large enough to capture seasonal variation of such features. In this context, Important Marine Mammal Areas, which are currently being discussed by the Marine Mammal Protected Areas Task Force and incorporate governmental and intergovernmental considerations (Corrigan et al., 2014), could help the delineation of sufficiently large protected areas. In addition, in a Marine Spatial Planning approach (Douvere, 2008), it would be worthwhile to overlay predicted density maps with anthropogenic pressure maps (Halpern et al., 2008) to define areas where pressures could be mitigated.

5. Conclusion

Habitat modelling of rare species is particularly challenging because habitat models require large datasets, yet rare species typically yield low numbers of sightings. As a result, combining datasets is a useful strategy to model the large-scale habitats of deep-divers; beaked whales, sperm whales and kogiids, across the North Atlantic Ocean and the Mediterranean Sea. At a local scale, predicted relative densities of deep-diving cetaceans were consistent with previous studies. At a larger scale, a gradient in predicted relative densities emerged, with the highest relative densities predicted on the western side of the study area. This pattern was evidenced thanks to assembling a large dataset and had not been detected previously. It highlighted the pronounced influence of active frontal zones, such as the Gulf Stream, on deep-diving cetaceans. Even though extensive gaps remain at a large scale, we were able to predict the habitats of these taxa throughout the North Atlantic Ocean and adjacent Mediterranean Sea, thus identifying potential habitats, including in non-sampled areas. However, these predictions should be used with caution as most of the study area represented geographical extrapolations and about half (mostly deeper waters) represented environmental extrapolations. Indeed, through an environmental space coverage gap analysis, we identified areas in tropical and deep oceanic waters where sampling effort was insufficient to predict habitats and needs to be increased to improve prediction reliability.

References

Anderson, D.R., & Burnham, K.P. (2002). Avoiding pitfalls when using information-theoretic methods. *The Journal of Wildlife Management* 66(3): 912–918.

Arcangeli, A., Campana, I., Marini, L., & MacLeod, C.D. (2015). Long-term presence and habitat use of Cuvier's beaked whale (*Ziphius cavirostris*) in the Central Tyrrhenian Sea. *Marine Ecology* 37: 269–282.

Austin, D., Bowen, W.D., McMillan, J.I., & Iverson, S.J. (2006). Linking Movement, Diving, and Habitat to Foraging Success in a Large Marine Predator. *Ecology* 87: 3095–3108.

Authier, M., Saraux, C., & Péron, C. (2016). Variable selection and accurate predictions in habitat modelling: a shrinkage approach. *Ecography* 40: 549–560.

Barlow, J. (2015). Inferring trackline detection probabilities, g (0), for cetaceans from apparent densities in different survey conditions. *Marine Mammal Science* 31(3): 923–943.

Bost, C.A., Cotté, C., Bailleul, F., Cherel, Y., Charrassin, J.B., Guinet, C., … Weimerskirch, H. (2009). The importance of oceanographic fronts to marine birds and mammals of the southern oceans. *Journal of Marine Systems* 78: 363–376.

Buckland, S.T., Anderson, D.R., Burnham, H.P., Laake, J.L., Borchers, D.L. & Thomas, L. (2001). Introduction to Distance Sampling: Estimating Abundance of Biological Populations. Oxford University Press, Oxford.

Buckland, S.T., Rexstad, E.A., Marques, T. A., & Oedekoven, C.S. (2015). *Distance sampling: methods and applications*. Springer.

Canada Meteorological Center. (2012). GHRSST Level 4 CMC0.2deg Global Foundation Sea Surface Temperature Analysis (GDS version 2). Ver. 2.0. PO.DAAC, CA, USA.

Cañadas, A., Sagarminaga, R., De Stephanis, R., Urquiola, E., & Hammond, P.S. (2005). Habitat preference modelling as a conservation tool: proposals for marine protected areas for cetaceans in southern Spanish waters. Aquatic Conservation: *Marine and Freshwater Ecosystems* 15(5): 495–521.

Cañadas, A., de Soto, N.A., Aissi, M., Arcangeli, A., Azzolin, M., B-Nagy, A., … Roger, Th. (2018). The challenge of habitat modelling for threatened low density species using heterogeneous data: The case of Cuvier's beaked whales in the Mediterranean. *Ecological Indicators* 85: 128–136.

Carrillo, M., & Ritter, F. (2010). Increasing numbers of ship strikes in the Canary Islands: proposals for immediate action to reduce risk of vessel-whale collisions. *Journal of Cetacean Research and Management* 11(2): 131–138.

Corrigan, C.M., Ardron, J.A., Comeros-Raynal, M.T., Hoyt, E., Notarbartolo Di Sciara, G., & Carpenter, K.E. (2014). Developing important marine mammal area criteria: learning from ecologically or biologically significant areas and key biodiversity areas. *Aquatic Conservation: Marine and Freshwater Ecosystems* 24(S2): 166–183.

D'Amico, A., Gisiner, R.C., Ketten, D.R., Hammock, J.A., Johnson, C., Tyack, P.L., & Mead, J. (2009). Beaked whale strandings and naval exercises. *Aquatic Mammals* 35: 452–472.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... Münkemüller, T. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36(1): 27–46.

Douvere, F. (2008). The importance of marine spatial planning in advancing ecosystem-based sea use management. *Marine policy* 32(5): 762–771.

ESRI, (2016). ArcGIS - A Complete Integrated System Environmental Systems Research Institute, Inc., Redlands, California. <http://esri.com/arcgis>.

Fernández, A., Edwards, J.F., Rodríguez, F., Espinosa de los Monteros, A., Herráez, P., Castro, P., … Arbelo, M. (2005). Gas and Fat Embolic Syndrome" Involving a Mass Stranding of Beaked Whales (Family *Ziphiidae*) Exposed to Anthropogenic Sonar Signals. *Veterinary Pathology* 42: 446–457.

Fleishman, E., Nally, R.M., Fay, J.P., & Murphy, D.D. (2001). Modeling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the Great Basin. *Conservation biology* 15(6): 1674–1685.

Foster, S.D., & Bravington, M.V. (2013). A Poisson-Gamma model for analysis of ecological non-negative continuous data. *Environmental and ecological statistics* 20: 533–552.

Hammond, P.S., Macleod, K., Berggren, P., Borchers, D.L., Burt, L., Cañadas, A., … Vázquez, J.A. (2013). Cetacean abundance and distribution in European Atlantic shelf waters to inform conservation and management. *Biological Conservation* 164: 107–122.

Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'agrosa, C., … Watson, R. (2008). A global map of human impact on marine ecosystems. *Science* 319(5865): 948–952.

Harris, P.T., Macmillan-Lawler, M., Rupp, J., & Baker, E.K. (2014). Geomorphology of the oceans. *Marine Geology* 352: 4–24.

Hastie, T., & Tibshirani, R. (1986). Generalized Additive Models. *Statistical Science* 3: 297–313.

Hazen, E.L., Palacios, D.M., Forney, K.A., Howell, E.A., Becker, E., Hoover, A.L., ... Bailey, H. (2017). WhaleWatch: a dynamic management tool for predicting blue whale density in the California Current. *Journal of Applied Ecology* 54(5): 1415–1428.

Hedley, S.L., & Buckland, S.T. (2004). Spatial models for line transect sampling. *Journal of Agricultural, Biological, and Environmental Statistics* 9(2): 181–199.

Jennings, M.D. (2000). Gap analysis : concepts, methods, and recent results. *Landscape Ecology* 15: 5–20.

Kinney, S.K., & Dunson, D.B. (2008). Bayesian model uncertainty in mixed effects models. *Random effect and latent variable model selection.* Springer New York. pp. 37-62.

King, G., & Zeng, L. (2007). When Can History be Our Guide ? The Pitfalls of Counterfactual Inference. *International Studies Quarterly* 51: 183–210.

Lanfredi, C., Azzellino, A., D'Amico, A., Centurioni, L., Ampolo Rella, M., Pavan, G., & Podestà, M. (2016). Key Oceanographic Characteristics of Cuvier's Beaked Whale (*Ziphius cavirostris*) Habitat in the Gulf of Genoa (Ligurian Sea, NW Mediterranean). *Journal of Oceanography and Marine Research* 4: 145.

Laran, S., Authier, M., Van Canneyt, O., Dorémus, G., Watremez, P., & Ridoux, V. (2017). A comprehensive survey of pelagic megafauna: their distribution, densities and taxonomic richness in the tropical Southwest Indian Ocean. *Frontiers in Marine Science* 4: 139.

Lumaret, J.P., & Jay-Robert, P. (2002). Modelling the species richness distribution of French dung beetles (*Coleoptera, Scarabaeidae*) and delimiting the predictive capacity of different groups of explanatory variables. *Global Ecology and Biogeography* 11(4): 265–277.

Madsen, P.T., de Soto, N.A., Tyack, P.L., & Johnson, M. (2014). Beaked whales. *Current Biology* 24(16): 728–730.

Mannocci, L., Catalogna, M., Dorémus, G., Laran, S., Lehodey, P., Massart, W., …& Ridoux, V. (2014). Predicting cetacean and seabird habitats across a productivity gradient in the South Pacific gyre. *Progress in Oceanography* 120: 383–398.

Mannocci, L., Roberts, J.J., Miller, D.L., & Halpin, P.N. (2017). Extrapolating cetacean densities to quantitatively assess human impacts on populations in the high seas. *Conservation Biology*. 31: 601–614.

Mannocci, L., Roberts, J.J., Halpin, P.N., Authier, M., Boisseau, O., Bradai, M.N., … Vella, J. (2018). Assessing cetacean surveys throughout the Mediterranean Sea: a gap analysis in environmental space. *Scientific Reports* 8(1): 3126.

Marques, F.F., & Buckland, S.T. (2003). Incorporating covariates into standard line transect analyses. *Biometrics* 59(4): 924-935.

McAlpine, D.F. (2009). *Pygmy and dwarf sperm whales.* Encyclopedia of marine mammals 2nd Edition. pp 936–938. Academic Press.

McShea, W.J. (2014). What Are the Roles of Species Distribution Models in Conservation Planning? *Environmental Conservation* 41: 93–96.

McSweeney, D.J., Baird, R.W., & Mahaffy, S.D. (2007). Site fidelity, associations, and movements of Cuvier's (*Ziphius cavirostris*) and Blainville's (*Mesoplodon densirostris*) beaked whales off the island of Hawai'i. *Marine Mammal Science* 23: 666–687.

Perrin, W.F., Würsig, B., & Thewissen, J.G.M. (Eds.). (2009). *Encyclopedia of marine mammals.* Academic Press.

Podestà, M., D'Amico, A., Pavan, G., Drougas, A., Komnenou, A., & Portunato, N. (2005). A review of Cuvier's beaked whale strandings in the Mediterranean Sea. *Journal of Cetacean Research and Management 7*(3): 251–261.

Plummer, M. (2016). rjags: Bayesian Graphical Models using MCMC. R package version 4-6. https://CRAN.R-project.org/package=rjags

Praca, E., & Gannier, A. (2007). Ecological niche of three teuthophageous odontocetes in the northwestern Mediterranean Sea. *Ocean Science Discussions* 4: 49–59.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Redfern, J.V., Moore, T.J., Fiedler, P.C., de Vos, A., Brownell, R.L., Forney, K.A., … Ballance, L.T. (2017). Predicting cetacean distributions in data-poor marine ecosystems. *Diversity and Distributions* 23: 394–408.

Roberts, J.J., Best, B.D., Dunn, D.C., Treml, E.A., & Halpin, P.N. (2010). Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environmental Modelling & Software* 25: 1197–1207.

Roberts, J.J., Best, B.D., Mannocci, L., Fujioka, E., Halpin, P.N., Palka, D.L., … Lockhart, G.G. (2016). Habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico. *Scientific Report* 6.

Rogan, E., Cañadas, A., Macleod, K., Santos, M.B., Mikkelsen, B., Uriarte, A., … Hammond, P.S. (2017). Distribution abundance and habitat use of deep diving cetaceans in the North-East Atlantic. Deep Sea Research Part II: Topical Studies in Oceanography 141: 8–19.

Royle, J.A., & Dorazio, R.M. (2008). *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities.* Academic Press.

Spiess, A. (2014). qpcR: Modelling and analysis of real-time PCR data. R package version 1.4-0. https://CRAN.R-project.org/package=qpcR

Spitz, J., Cherel, Y., Bertin, S., Kiszka, J., Dewez, A., & Ridoux, V. (2011). Prey preferences among the community of deep-diving odontocetes from the Bay of Biscay, Northeast Atlantic. *Deep Sea Research I* 58: 273–282.

Staudinger, M.D., McAlarney, R.J., McLellan, W.A., & Ann Pabst, D. (2014). Foraging ecology and niche overlap in pygmy (*Kogia breviceps*) and dwarf (*Kogia sima*) sperm whales from waters of the US mid-Atlantic coast. *Marine Mammal Science* 30(2): 626–655.

Stoll, H., King, G., & Zeng, L. (2014). WhatIf: Software for Evaluating Counterfactuals. R package version 1.5-6. https://cran.r-project.org/web/packages/WhatIf/index.html.

Stone, C.J., & Tasker, M.L. (2006). The effects of seismic airguns on cetaceans in UK waters. *Journal of Cetacean Research and Management* 8: 255–263.

Thomas, L., Buckland, S.T., Rexstad, E.A., Laake, J.L., Strindberg, S., Hedley, S.L., … Burnham, K.P. (2010). Distance software: Design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology* 47: 5–14.

Unger, B., Rebolledo, E.L.B., Deaville, R., Gröne, A., IJsseldijk, L.L., Leopold, M.F., … Herr, H. (2016). Large amounts of marine debris found in sperm whales stranded along the North Sea coast in early 2016. *Marine pollution bulletin* 112(1): 134–141.

Virgili, A., Authier, M., Monestiez, P., & Ridoux, V. (2018). How many sightings to model rare marine species distributions. *PloS one* 13(3): e0193231.

Waring, G.T., Hamazaki, T., Sheehan, D., Wood, G., & Baker, S. (2001). Characterization of beaked whale (*Ziphiidae*) and sperm whale (*Physeter macrocephalus*) summer habitat in shelf-edge and deeper waters off the Northeast U.S. *Marine Mammal Science* 17: 703–717.

Wimmer, T., & Whitehead, H. (2004). Movements and distribution of northern bottlenose whales, *Hyperoodon ampullatus*, on the Scotian Slope and in adjacent waters. *Canadian Journal of Zoology* 82: 1782–1794.

Wood, S.N. (2006). On confidence intervals for generalized additive models based on penalized regression splines. *Australian & New Zealand Journal of Statistics* 48: 445–464.

Wood, S. (2013). mgcv: Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation. Retrieved 7 July 2014, from http://cran.r-project.org/web/packages/mgcv/index.html.

Woodson, C.B., & Litvin, S.Y. (2015). Ocean fronts drive marine fishery production and biogeochemical cycling. *Proceedings of the National Academy of Sciences* 112(6): 1710–1715.

DATA ACCESSIBILITY

All sighting and effort data used in this study are available in the OBIS SEAMAP database: http://seamap.env.duke.edu/. All data providers can be contacted via the OBIS SEAMAP website.

Tables

Table 1. Effort performed by platform type or Beaufort sea-state for all surveys in the North Atlantic Ocean and the Mediterranean Sea. This table presents the total effort conducted in each sector broken down by platform type and Beaufort sea-state. Beaufort sea-state values reported with decimals in the surveys were rounded up. For the analyses, all segments with Beaufort sea-state > 4 were excluded. 'NE-ATL' means northeast Atlantic Ocean; 'NW-ATL' means northwest Atlantic Ocean and 'MED' means Mediterranean Sea.

Sectors	Total survey effort (km and %)	Total aerial effort (km)	Total shipboard effort (km)	Total effort by Beaufort sea-state class (km)				
				0-1	1-2	2-3	3-4	4-7
NE-ATL	469,000 37 %	70,000	399,000	77,000	118,000	136,000	85,000	53,000
NW-ATL	557,000 45 %	546,000	11,000	43,000	121,000	199,000	132,000	62,000
MED	195,000 16 %	87,000	109,000	92,000	70,000	27,000	6,000	800
TROPICS	19,000 2 %	15,000	4,000	11,000	3,000	4,000	2,000	400
STUDY AREA	1,240,000	718,000 58 %	522,000 42 %	222,000 18 %	312,000 25 %	365,000 30 %	225,000 18 %	116,000 9%

Table 2. Candidate environmental predictors used for the habitat modelling. All variables were resampled at a 0.25° resolution. A: Depth and slope were derived from GEBCO-08 30 arc-second database (http://www.gebco.net/); 30 arc-second is approximately equal to 0.008°. B: Surface area per cell was calculated in ArcGIS 10.3 from the shapefile of canyons and seamounts provided by Harris et al. (2014). C: The mean, standard error and gradient of Sea Surface Temperature (SST) were calculated from the GHRSST Level 4 CMC SST v.2.0 (Canada Meteorological Centre, https://podaac.jpl.nasa.gov/dataset/CMC0.2deg-CMC-L4-GLOB-v2.0).

D: The Aviso ¼° DT-MADT geostrophic currents dataset was used to compute mean and standard deviation of Sea Surface Height (SSH) and Eddy Kinetic Energy (EKE; https://www.aviso.altimetry.fr/en/data/products/sea-surface-height-products/global/madt-h-uv.html). E: Net primary production (NPP) was derived from SeaWIFS and Aqua using the Vertically Generalised Production Model (VGPM; http://orca.science.oregonstate.edu/1080.by.2160.8day.hdf.vgpm.m.chl.m.sst.php).

Environmental variable	Original Resolution	Source	Justification
Physiographic			
Depth (m)	30 arc sec	A	Deep-divers feed on squids and fish in the deep water column
Slope (°)	30 sec arc	A	Associated with currents, high slope induce prey aggregation or enhanced primary production
Surface area of canyons and seamounts in a 0.25° cell (km²)	30 sec arc	B	Deep-divers are often associated with canyons and seamounts structures; the variable indicates the proportion of this habitat in each cell
Oceanographic			
Mean of SST (°C)	0.2°, daily	C	Variability over time and horizontal gradients of SST reveal front locations, potentially associated with prey aggregations or enhanced primary production
Standard error of SST (°C)	0.2°, daily	C	
Mean gradient of SST (°C)	0.2°, daily	C	
Mean of SSH (m)	0.25°, daily	D	High SSH is associated with high mesoscale activity and enhanced prey aggregation or primary production
Standard deviation of SSH (m)	0.25°, daily	D	
Mean of EKE ($m^2.s^{-2}$)	0.25°, daily	D	High EKE relates to the development of eddies and sediment resuspension induce prey aggregation
Standard error of EKE ($m^2.s^{-2}$)	0.25°, daily	D	
Mean of NPP ($mgC.m^{-2}.day^{-1}$)	9 km, 8 days	E	Net primary production as a proxy of prey availability

Table 3. Encounter rates in sightings·100 km^{-1} calculated for the entire study area and each sub-region of the North Atlantic Ocean and the Mediterranean Sea. 'NE-ATL' means northeast Atlantic Ocean; 'NW-ATL' means northwest Atlantic Ocean and 'MED' means Mediterranean Sea.

	NE-ATL	NW-ATL	MED	TROPICS	STUDY AREA
Beaked whales	0.042	0.058	0.035	0.22	0.051
Sperm whales	0.057	0.067	0.09	0.095	0.067
Kogiids	0.0013	0.01	0.0	0.23	0.0085

Table 4. Summary of the selected models by species group.

Species group	Selected variables	Explained deviance	Akaike weight	Specific comments
Beaked whale	Depth Gradients SST Slope NPP	33.1 %	0.98	Depth, gradients SST and slope selected in the first 10 models
Sperm whale	Depth Gradients SST SSH mean NPP	20.6 %	0.76	Depth, gradients SST and SSH mean selected in the first 8 models
Kogiids	Depth Gradients SST EKE mean Surface of canyons and seamounts	55.7 %	0.17	Depth, gradients SST and surface of canyons and seamounts selected in the first 7 models

Figures

Fig. 1. Study area divided into sub-regions showing assembled survey effort (A), along with the beaked whale (B), sperm whale (C) and kogiid (D) sightings recorded during all surveys. The blue polygon delineates overall study area and other polygons delineate sub-regions. Surveys were carried out along transects following a line-transect methodology (survey details in Appendix S1 in Supporting Information). Sightings were classified by group sizes with each point representing one group of individuals and point size representing the number of animals in a group.

Fig. 2. Beaked whale and sperm whale averaged ESWs estimated for each survey group and each platform type. For each survey group, the boxplot represents the extent of estimated ESWs depending on Beaufort sea-states and observation heights recorded within the group.

Fig. 3. Functional relationships for the selected variable (A) and the predicted relative densities of beaked whales in individuals·km^{-2} (B and C). A: Solid lines are the estimated smooth functions, and the shaded regions represent the approximate 95% confidence intervals. The y-axes indicate the number of individuals on a log scale, where zero indicates no effect of the covariate. The vertical lines indicate the 2.5th and 97.5th quantiles of the data. Black areas on prediction maps (B: without precautionary approach and C: with a 5% precautionary approach) represent zones where we did not extrapolate the predictions. Percentages represent the proportion of the study area defined as interpolation with the gap analysis.

Fig. 4. Functional relationships for the selected variable (A) and the predicted relative densities of sperm whales in individuals·km^{-2} (B and C). A: Solid lines are the estimated smooth functions, and the shaded regions represent the approximate 95% confidence intervals. The y-axes indicate the number of individuals on a log scale, where zero indicates no effect of the covariate. The vertical lines indicate the 2.5th and 97.5th quantiles of the data. Black areas on prediction maps (B: without precautionary approach and C: with a 5% precautionary approach) represent zones where we did not extrapolate the predictions. Percentages represent the proportion of the study area defined as interpolation with the gap analysis.

Fig. 5. Functional relationships for the selected variable (A) and the predicted relative densities of kogiids in individuals·km^{-2} (B and C). A: Solid lines are the estimated smooth functions, and the shaded regions represent the approximate 95% confidence intervals. The y-axes indicate the number of individuals on a log scale, where zero indicates no effect of the covariate. The vertical lines indicate the 2.5th and 97.5th quantiles of the data. Black areas on prediction maps (B: without precautionary approach and C: with a 5% precautionary approach) represent zones where we did not extrapolate the predictions. Percentages represent the proportion of the study area defined as interpolation with the gap analysis.

Supporting Information

Appendix S1: Characteristics of the study area.

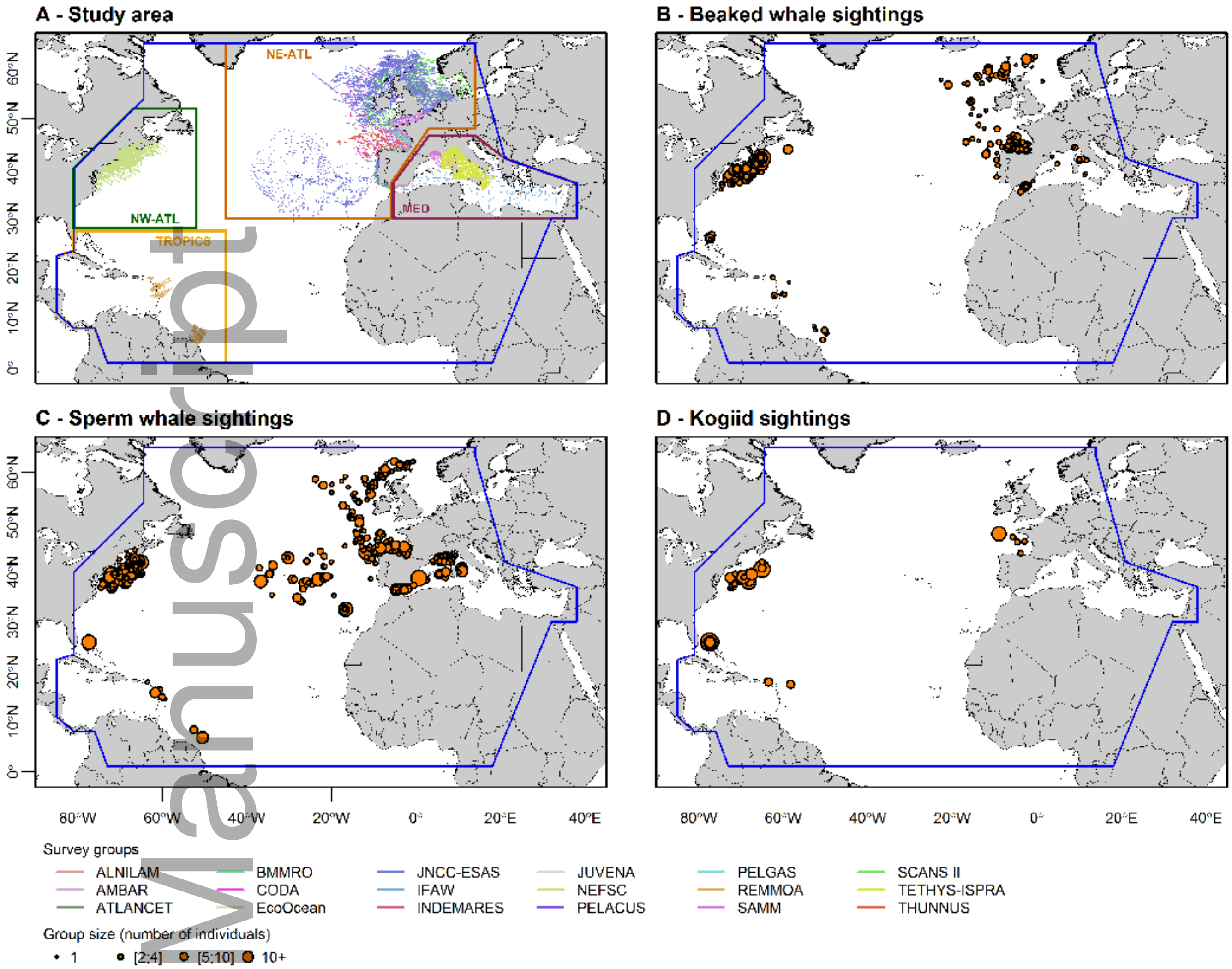
Appendix S2: Details of surveys used in the analyses. Total effort represents the total length of transects of each survey (without removing the transects with a Beaufort sea-state > 4). MED: Mediterranean Sea; NE-ATL: Northeast Atlantic Ocean; NW-ATL: Northwest Atlantic Ocean.

Appendix S3: Monthly environmental conditions averaged over the study period (from 1998 to 2015).

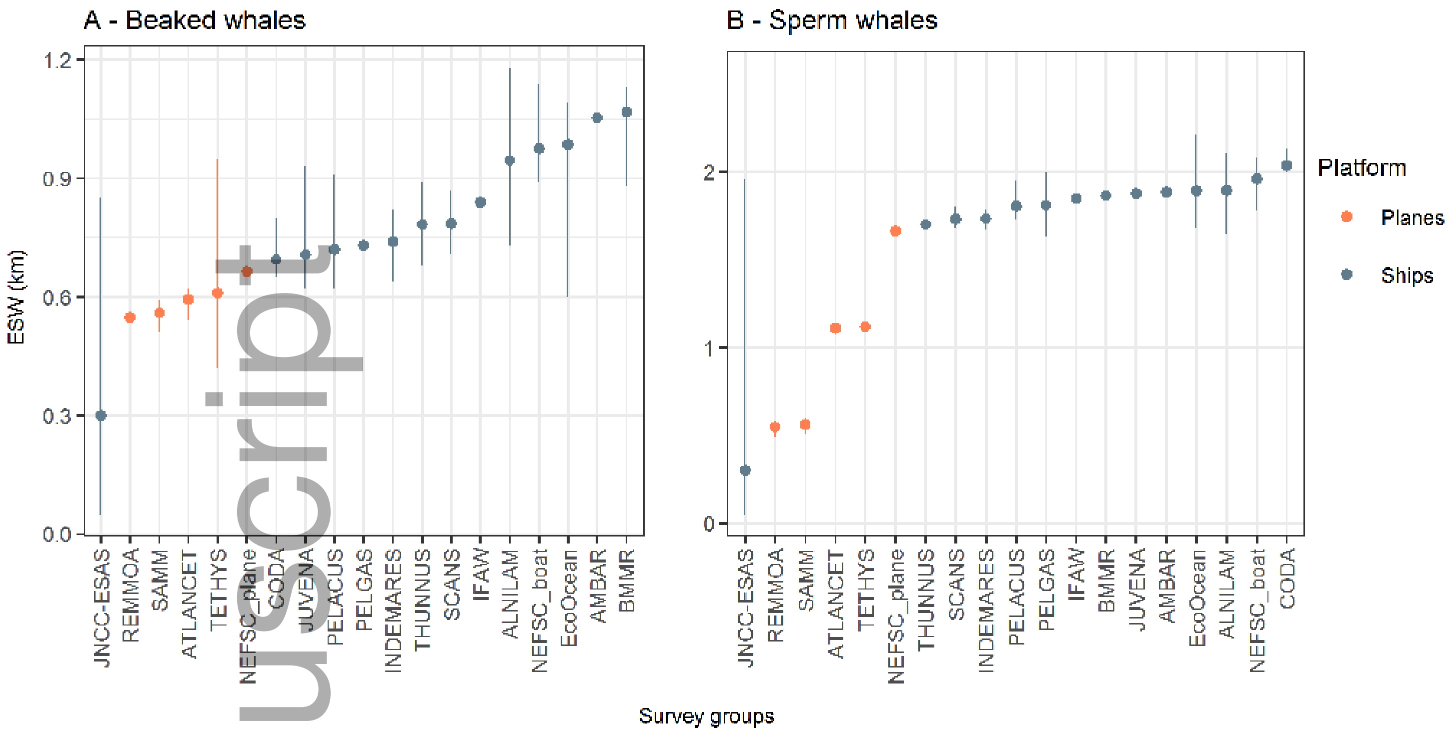
Appendix S4: Effective Strip Width estimation methodology.

Appendix S5: Supporting information for the models.

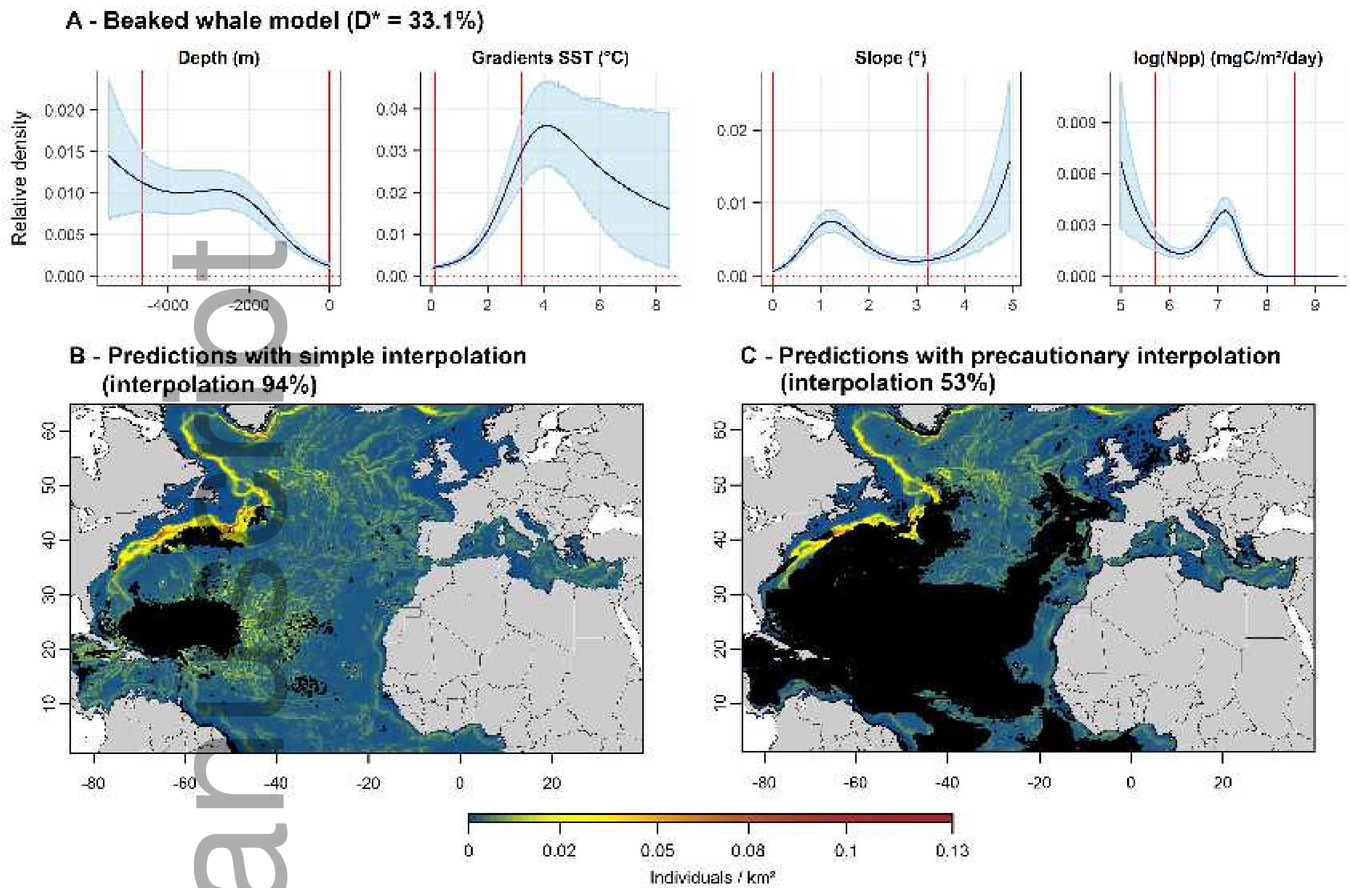
Appendix S6: Comparison between predictions of the 5 best kogiid models, the average prediction of the 5 best models (Mean) and the average prediction of the 5 best models weighted by the Akaike weight (Weighted). The 5 models are described in the table at the bottom (mod: model; AIC: Akaike Information Criterion). "Mean" is the simple average of the predictions of the 5 best models. To calculate the "Weighted" prediction, we averaged the predictions of the 5 best models by weighting each prediction by the Akaike weight (weighted.mean function of the raster package).



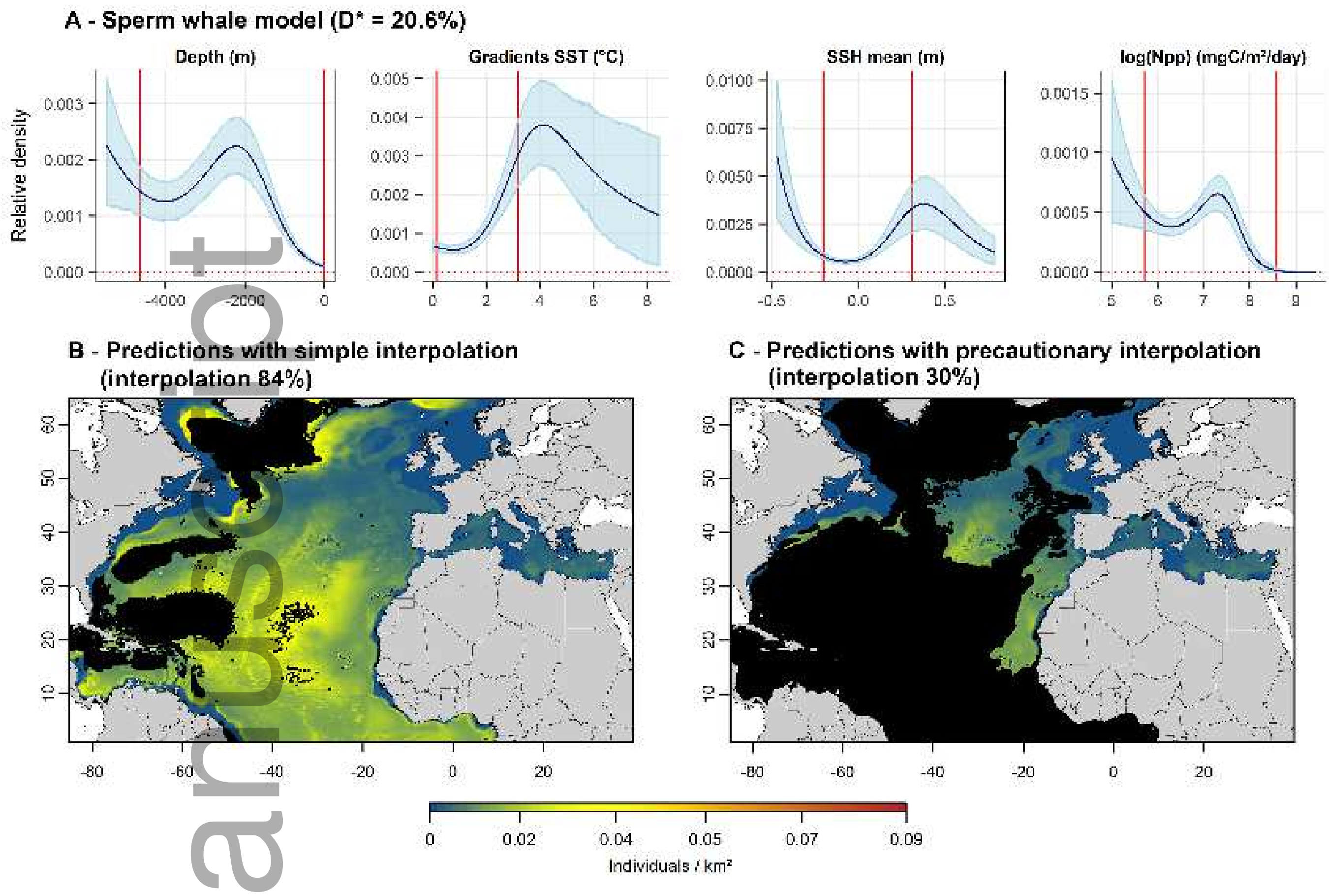
geb_12850_f1.tif



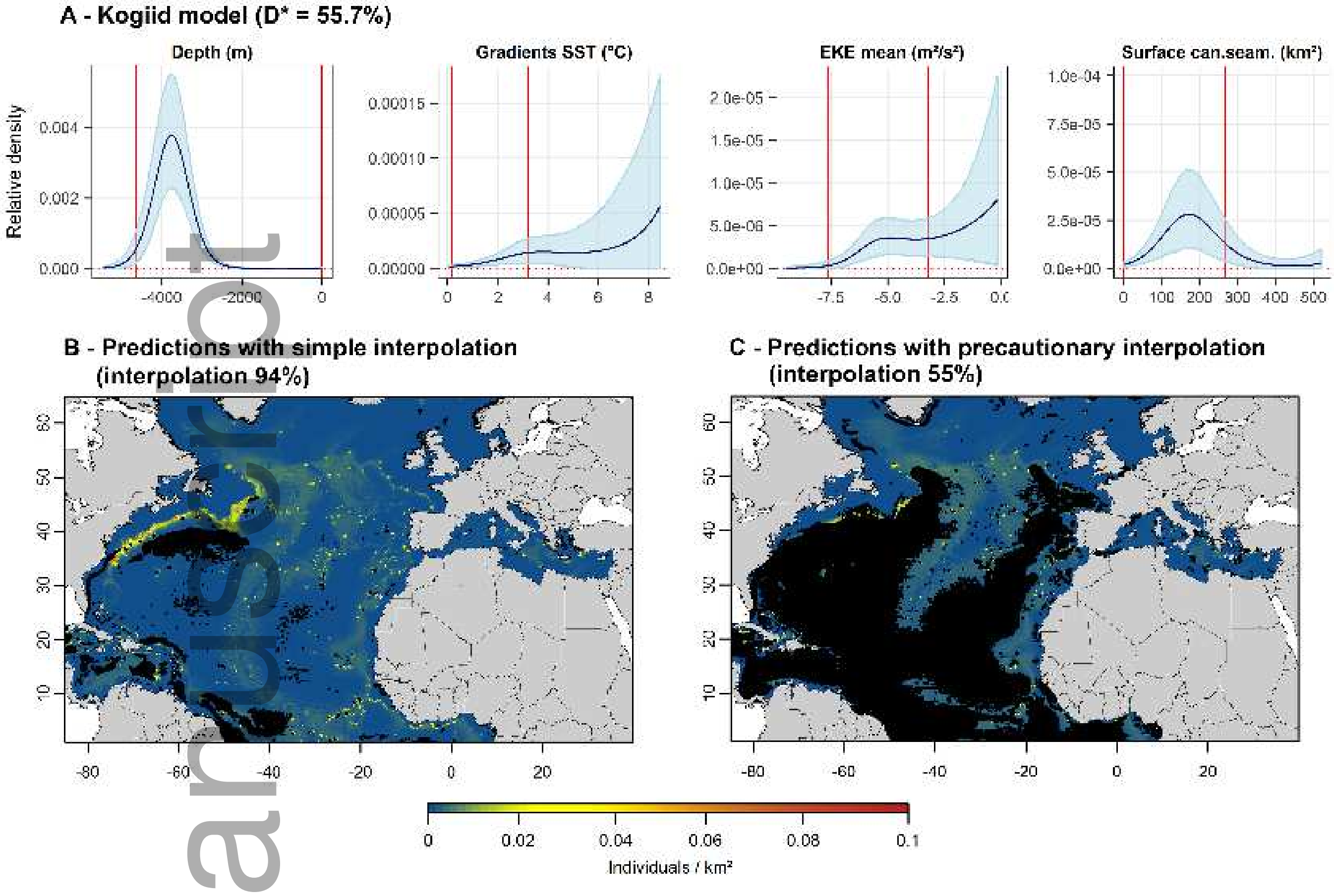
geb_12850_f2.tiff



geb_12850_f3.tif



geb_12850_f4.tif



geb_12850_f5.tif