Journal Pre-proof

Assessing the reliability of species distribution models under changing environments: a case study on cetaceans in the North-East Atlantic

Rémi Pigeault, Matthieu Authier, Nadya C. Ramirez-Martinez, Auriane Virgili, Steve C.V. Geelhoed, Jan Haelters, Maite Louzao, Camilo Saaveedra, Anita Gilles

PII: S2351-9894(24)00503-1

DOI: https://doi.org/10.1016/j.gecco.2024.e03299

Reference: GECCO3299

To appear in: *Global Ecology and Conservation*

Received date: 7 August 2024
Revised date: 25 October 2024
Accepted date: 8 November 2024

Please cite this article as: Rémi Pigeault, Matthieu Authier, Nadya C. Ramirez-Martinez, Auriane Virgili, Steve C.V. Geelhoed, Jan Haelters, Maite Louzao, Camilo Saaveedra and Anita Gilles, Assessing the reliability of species distribution models under changing environments: a case study on cetaceans in the North-East Atlantic, *Global Ecology and Conservation,* (2024) doi:https://doi.org/10.1016/j.gecco.2024.e03299

Assessing the reliability of species distribution models under changing environments: a case study on cetaceans in the North-East Atlantic

Rémi Pigeault[a], Matthieu Authier[b], Nadya C. Ramirez-Martinez[a,c], Auriane Virgili[b,1], Steve C. V. Geelhoed[d], Jan Haelters[e], Maite Louzao[f], Camilo Saaveedra[g], Anita Gilles[a*]

[a]*Institute for Terrestrial and Aquatic Wildlife Research, University of Veterinary Medicine Hannover, Foundation, Buesum, Germany*

[b]*Observatoire PELAGIS, UAR 3462, La Rochelle Université/CNRS, La Rochelle, France*

[c]*Fundación Macuáticos Colombia, Calle 27 # 79-167, Medellin, Antioquia, Colombia*

[d]*Wageningen Marine Research, Den Helder, the Netherlands*

[e]*Institute of Natural Sciences, Ostend, Belgium*

[f]*AZTI, Marine Research, Basque Research and Technology Alliance (BRTA), Herrera Kaia Portualdea z/g, 20110 Pasaia, Spain*

[g]*Oceanographic Centre of Vigo, Spanish Institute of Oceanography, Spanish National Research Council (IEO-CSIC), Vigo, Spain*

[1]*Present address: Share the Ocean, Larmor Baden, France*

*Correspondence: Anita Gilles, Institute for Terrestrial and Aquatic Wildlife Research, University of Veterinary Medicine Hannover, Foundation, Werftstraße 6, D-25761 Buesum, Germany (e-mail: anita.gilles@tiho-hannover.de).

Abstract

1. Species Distribution Models (SDMs) assume stable relationships between species and their environment from which predictions are made. These relationships are likely to vary with changing environments, and predictions might depend more on modelling choices than on empirical data. Reliability assessments of predictions are necessary to support policy-making.
2. We identified environmental extrapolations among potential predictions of cetaceans' distribution from 2005 to 2020 in the North-East Atlantic and calculated the percentages of calibration data with similar environments (nearby data), supporting these predictions. Thus, the assessment of reliability is generic, as evaluated before model fitting.
3. Predictions on continental shelves were extensively supported by the calibration data and were more reliable throughout the year than predictions on continental slopes and abyssal plains, which were more supported in summer. Predictions off Portugal were particularly uncertain due to the lack of surveys in this region of deep, warmer waters with seamounts.
4. The high effort between May and July led to a southern winter shift of nearby data, following the decrease in temperatures. A large part of the predictions between December and April was extrapolated due to the low coverage of the winter primary productivity drops, spring peaks and cold waters. They were based on data collected during other seasons and regions, and given the large spatial extent of the area, and the seasonality and regionality of the cetacean distributions, reliable winter predictions might be restricted to geographic areas where winter surveys took place. These predictions are more uncertain and warrant caution.

5. *Synthesis and applications*: extrapolations and nearby data highlighted environmental gaps to predict cetacean distributions in the North-East Atlantic, which could be covered by future surveys. This informs model users of regions and periods when predictions reliability becomes uncertain. SDMs are invaluable tools for supporting conservation applications and, despite the warnings that have been issued, the degree of information available for predicting distribution is still rarely reported. We recommend adding this assessment as routine information on the reliability of predictions.

Keywords:

climate change, decision-making, environmental similarity, extrapolation, nearby data, robustness, transferability

1. Introduction

By relating species attributes such as presence or abundance to environmental covariates, Species Distribution Models (SDMs) are used to estimate species distributions, delimit protected areas or predict climate change impacts (Elith and Leathwick, 2009; Guisan et al., 2013; Redfern et al., 2006; Zurell et al., 2020). The reliance of SDMs on calibration data and modelling choices implies uncertainties and potential errors in predictions (e.g. quality and bias of calibration data, choice and source of environmental variables, modelling procedure; see Rocchini et al., 2011). Among these, the reliability of predictions from SDM decreases in environments different from those of the calibration dataset, as the fitted species-environment relationships are uncertain and may no longer be valid (Dormann, 2011; Elith and Leathwick, 2009; Fitzpatrick and Hargrove, 2009; Owens et al., 2013; Qiao et al., 2019). Although acknowledged, assessments of the environmental similarity between predictions and calibration dataset are still rarely reported along with predictions (Rousseau and Betts, 2022; Taheri et al., 2021; Yates et al., 2018). Spatially explicit maps of uncertainty are useful, in this regard, to depict the reliability of SDMs (*e.g.* as ignorance maps, see Tessarolo et al., 2021; Rocchini et al., 2011).

1.1. Assessing the SDM reliability in new environments

The environmental similarity provides an indication of the amount of data available to predict a species' distribution *a priori* (ahead of model fitting; Bouchet et al., 2020; King and Zeng, 2007; Mesgaran et al., 2014). If the environmental similarity between the calibration data and a prediction is low, the prediction depends more on modelling choices than on empirical data, and may not be reliable. Environmental extrapolations and nearby data have emerged as practical metrics to assess this environmental similarity and map the uncertainties (Bouchet et al., 2019; García-Barón et al., 2019; Mannocci et al., 2018; Virgili et al., 2017, 2019; Zurell et al., 2020).

The environment can be defined as the combination of the covariates (Hutchinson, 1957; Fitzpatrick and Hargrove, 2009) and the environmental space as the *n*-dimensional space where each dimension is an environmental covariate that describes the species' ecological niche (Hutchinson, 1957). Mesgaran et al. (2014) delimited the environmental coverage of the calibration dataset by the smallest hull that contains these environments in the environmental space (see Fig. 1 with an illustrative dataset and two covariates). Extrapolations were then defined as predictions with combinations of covariates outside the environmental coverage of calibration in the environmental space (Mesgaran et al., 2014), including, therefore, predictions outside the sampled range of one or more of the covariates. By contrast, interpolations were predictions with combinations of covariates inside the environmental coverage of calibration. Theoretically, the fitted species-environment relationships are expected to be more uncertain (and more dependent on modelling choices) outside the environmental coverage, so extrapolations are, in themselves, more uncertain than interpolations (Fitzpatrick and Hargrove, 2009; Mesgaran et al., 2014; Zurell et al., 2012). On the other hand, the more calibration data that have environments similar to a prediction (aka nearby data, Fig. 1), the better the prediction is data-driven, and hence more reliable (regardless of whether it is an interpolation or extrapolation; Bouchet et al., 2019; King and Zeng, 2007). Otherwise, the prediction depends more on modelling choices than on empirical evidence and is more uncertain. Both extrapolations and nearby data identify environmental gaps in the calibration dataset that future data collection (e.g. dedicated surveys) can fill.

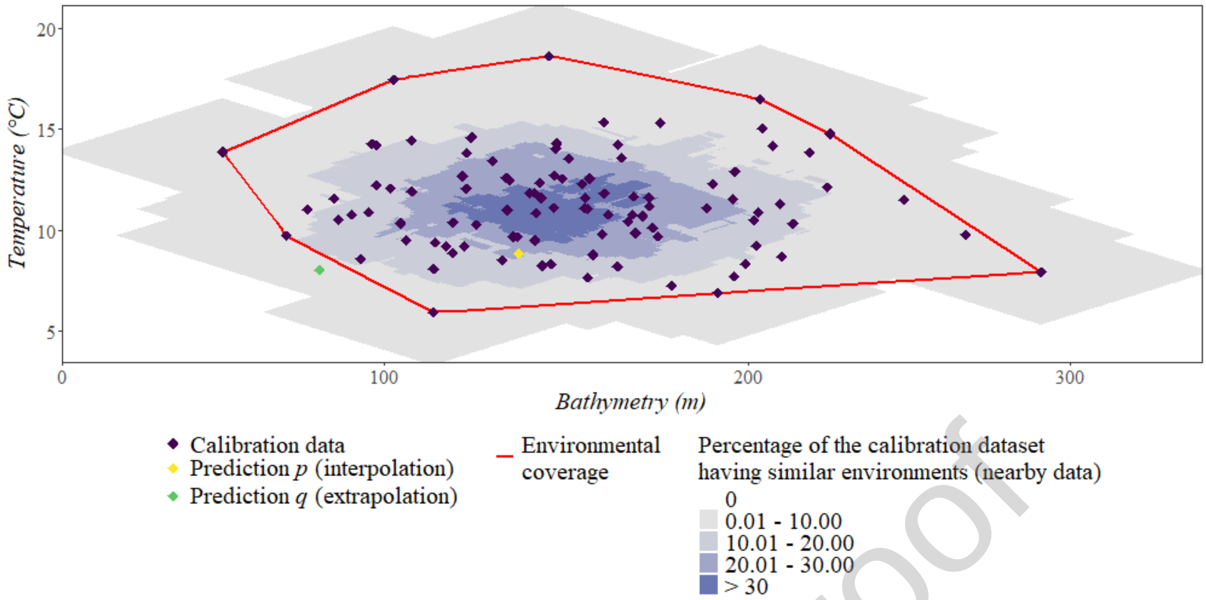


Figure 1. Representation of the environmental space covered by an illustrative dataset with two environmental covariates, temperature and bathymetry. Extrapolations (prediction *q*) are predictions under environments outside the environmental coverage (red line) in the environmental space, and interpolations (prediction *p*) are predictions inside the environmental coverage. The more nearby data a prediction has in the environmental space, the more this prediction is based on empirical data and is reliable. This figure results from a simulation using Gower's distance in the *WhatIf* R package (King and Zeng, 2007; version 1.5-8). Unlike Euclidean distances, the radius around a Gower's distance produces a diamond shape (distances in each dimension are not squared before being added together; see Methods).

1.2. Practical application: predictive reliability of SDMs for cetaceans in the North-East Atlantic

As top predators, cetaceans are key indicators of environmental status and can serve as prime sentinels of multitrophic marine ecosystem changes (Kiszka et al., 2015; Moore, 2008). They are protected by national and international legislation (e.g., EU Habitats Directive, Convention on Migratory Species and its regional agreements). In recent decades, large-scale international and national surveys have collected data on the occurrence of cetaceans in offshore and shelf waters of the North-East Atlantic (Hammond et al., 2002, 2013, 2021 ; CODA, 2009; Rogan et al., 2018; Gilles et al., 2016; Laran et al., 2017). Using data collected from surveys conducted between 2005 and 2020 (all seasons included) and a common set of environmental covariates, we identified environmental extrapolations among potential SDMs predictions in the North-East Atlantic from 2005 to 2020 and calculated their percentages of nearby data. Thus, we investigated the potential of the collected data to build reliable distribution models in the region. This assessment was performed for the indicator *Abundance and distribution of cetaceans* of the 2023 Quality Status Report (Geelhoed et al., 2022) of OSPAR (Convention for the Protection of the Marine Environment of the North-East Atlantic).

2. Methods

The process of collecting and processing survey data to create the calibration dataset, the prediction grids within the North-East Atlantic from 2005 to 2020 and the assessment of the predictive reliability is shown in Fig. 2 and detailed in the following sections.

2.1.Survey data & field methodology

In the North-East Atlantic, large-scale international and national surveys have been conducted to assess cetacean distributions and abundances over three decades using dedicated aerial and ship-based surveys. The survey data used in this analysis are those collated for the OSPAR 2023 Quality Status Report (Geelhoed et al., 2022). Large-scale surveys included SCANS-II (Small Cetaceans in the European Atlantic and North Sea; 2005; Hammond et al., 2013), SCANS-III (2016; Hammond et al., 2021), CODA (Cetacean Offshore Distribution and Abundance in the European Atlantic; CODA, 2009) and ObSERVE (Rogan et al., 2018) surveys. All surveys followed a distance-sampling methodology, i.e., allowing to estimate the detection probability from the transect line and to estimate absolute abundances (Buckland et al., 2009). Detailed descriptions of shipboard and aerial survey field methodologies are provided in previous studies (Gilles et al., 2016; Hammond et al., 2021; Scheidat et al., 2008).

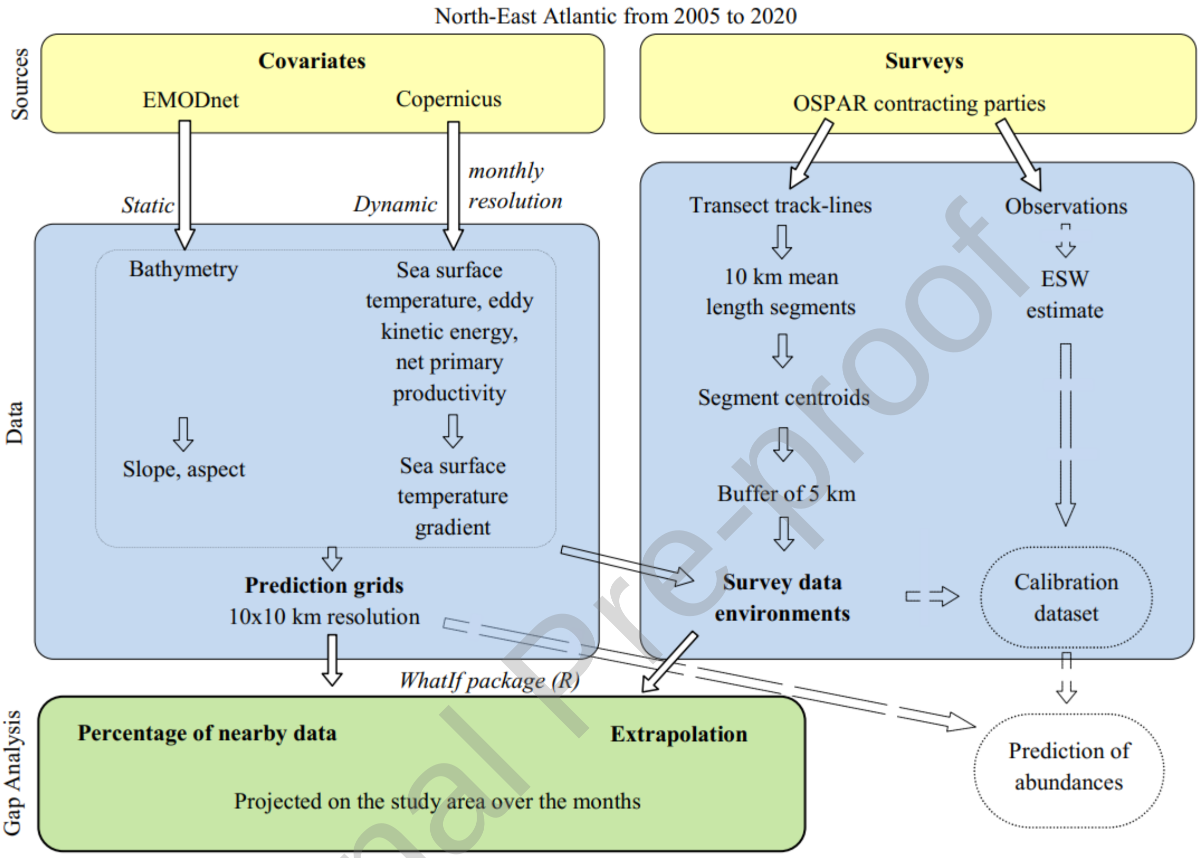


Figure 2. Flowchart of data preparation steps and analyses. The dashed arrows represent additional steps in the prediction of abundances. These steps were not taken in the present investigation.

2.2.Environmental covariates

The candidate set of covariates (Table 1) was selected following previous studies on cetacean distribution modelling (Gilles et al., 2016; Lambert et al., 2017; Rogan et al., 2017). These covariates are proxies for habitat suitability and prey distributions, probably the main drivers of cetacean distribution (Palacios et al., 2006, 2013; Redfern et al., 2006). Seven covariates were selected, at a monthly resolution, to limit the complexity of the model. Increasing the number of covariates increases the number of new possible combinations of covariates, and therefore the number of extrapolations (Authier et al., 2017).

2.3.Data preparation

Survey data collected from 2005 to 2020 were compiled, all seasons included, for the indicator *Abundance and distribution of cetaceans* in the OSPAR 2023 Quality Status Report (Geelhoed et al. 2022). The calibration data were the compiled data, clipped to the study area (Fig. 3). The study area encompassed the OSPAR regions II, III, and IV completed by the overlapping MSFD (EU Marine Strategy Framework Directive) sub-regions of the Celtic Seas, Bay of Biscay, Iberian Coast, and Greater North Sea, and covers a total surface of 2.46 10^6 km². The calibration data were segmented into 10 km mean length segments in the R software environment (R core team 2020), conforming with Becker et al. (2020), Gilles et al. (2016) and Virgili et al. (2019). The monthly covariate means were extracted within a radius of 5 km around their centroid. Monthly prediction grids were created, composed of 10x10 km cells, from January 2005 to December 2020. The monthly covariate means were averaged within each cell.

Table 1. Candidate environmental covariates used to model the cetacean distributions. Source A: EMODnet Bathymetry 2020 Digital Terrain Model (https://www.emodnet-bathymetry.eu/). Slope and aspect were derived from bathymetry using the *Terrain* function (*Raster* R-package). Source B: Copernicus (https://resources.marine.copernicus.eu): Global Ocean Physics Reanalysis (B_1), Global Ocean Biogeochemistry Hindcast (B_2; see Acknowledgements for DOIs). The SST gradients were calculated from the SST means, using the *DetectFronts* function (*Grec* R package).

Environmental variable		Original spatial resolution	Original temporal resolution	Temporal prediction resolution	Abbreviation	Source	Justification
Physiographic	Bathymetry (m)	1/16 arc minute	NA	NA	Bathy	A	Proxy for cetacean prey distribution (e.g., the deep-divers feed on squid and fish in the deep-water column).
	Slope (rad)				Slope		Associated with currents, high slopes induce enhanced primary production or prey aggregation.
	Aspect (rad)				Aspect		Describe currents and prominent structures such as canyons, seamounts or mountain chains, used as proxies for predator hotspots and useful in locations where access to biological data is limited.
Oceanographic	Sea surface temperature mean (°C) and gradient (°C/m)	0.083° (-5500 m to 0 m)	Monthly	Monthly	mSST, gradSST	B_1	Variability over time and horizontal gradients of SST reveal front locations, mixing of water and is associated with enhanced primary production and prey aggregations.
	Eddy kinetic Energy (m/s)				EKE		High EKE relates to the development of eddies, upwelling of nutrients and enhanced primary production, which induce prey aggregation.
	Net primary productivity ($mg.m^{-3}.day^{-1}$)	0.25° (-5500 m to 0 m)			NPPV	B_2	Net primary production is a proxy of zooplankton distribution, feeding cetacean prey.

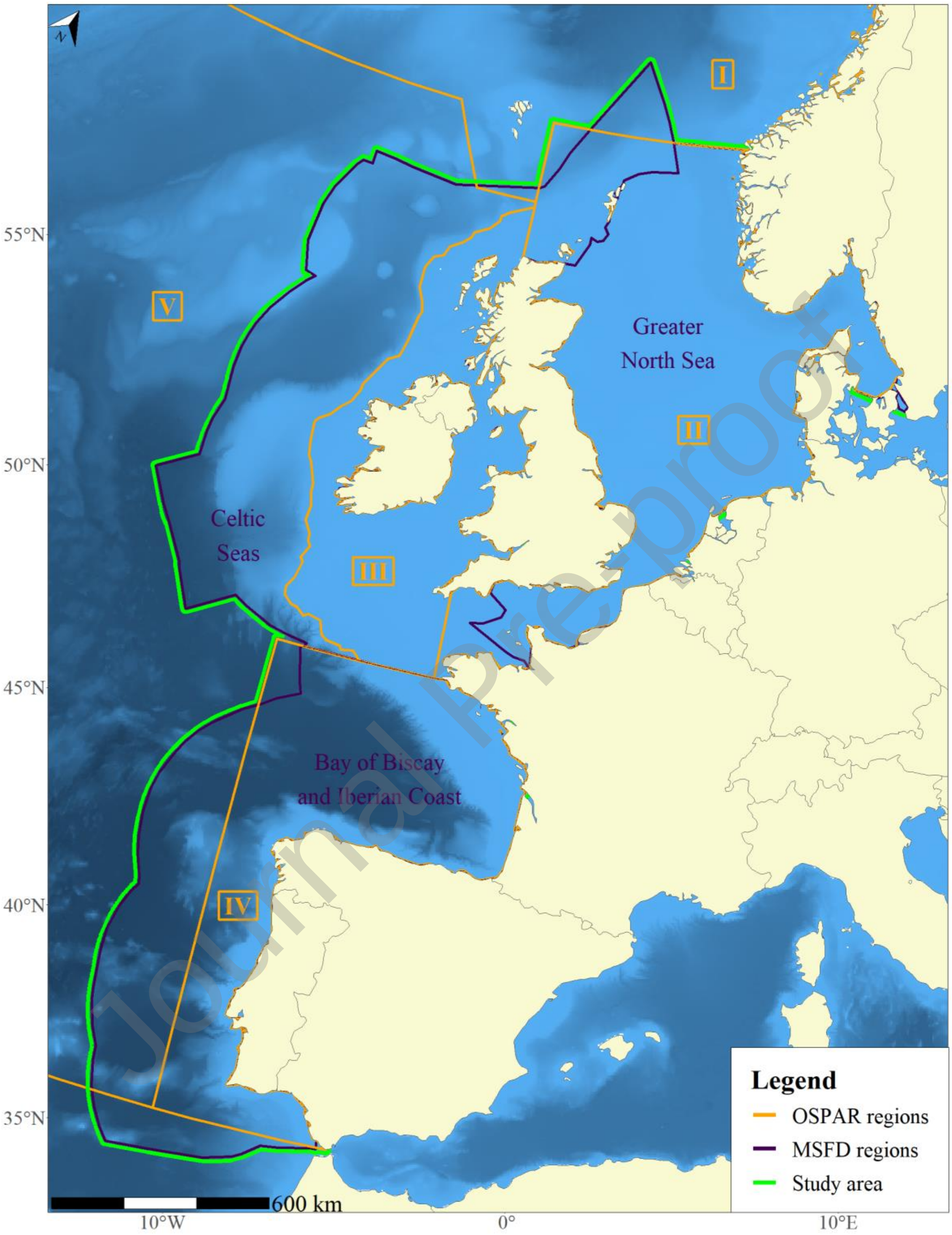


Figure 3. Study area delimited by OSPAR regions II, III and IV and the overlapping MSFD sub-regions (EU Marine Strategy Framework Directive). Bathymetry source: EMODnet DTM (https://www.emodnet-bathymetry.eu/).

2.4.Extrapolations and nearby data

The method followed is that of Bouchet et al. (2019). Nonetheless, *dsmextra* R package (Bouchet et al., 2020) could not be run due to an undetermined error related to the grid irregularity. The extrapolations and percentages of nearby data of the predictions from 2005 to 2020 were therefore estimated with the *WhatIf* R-package version 1.5-8 (King and Zeng, 2007), using the aggregated dataset including all seasons. Extrapolations are binary data, defining whether the prediction environment falls outside or inside the environmental coverage of the calibration dataset in the environmental space (Fig. 1). Nearby data for a prediction are calibration data located in the environmental space within one mean geometric Gower's distance (mean Gower's distance between all possible pairs of calibration data). The Gower's distance G^2 between two points, i and j (here, all possible pairs of calibration data), is defined as the average absolute distance between the values of these two points in each dimension, divided by the dimension range. With K environmental dimensions, the Gower's formula is as follows:

$$G_{ij}^2 = \frac{1}{K}\sum_{k=1}^{K}\frac{|x_{ik} - x_{jk}|}{r_k}$$

Here, r_k is the difference between the highest and lowest values of the k^{th} covariate in the dataset. By equalizing the variances of the covariates, they contribute equally to the Gower's distance. Nearby data for a prediction are calibration data that have similar environmental conditions.

3. Results

3.1.Spatio-temporal coverage of calibration

The dataset consisted of 74,430 points, totalling 698,551 km of survey effort between 2005 and 2020. The most surveyed years were 2005, 2012 and 2016, and the best covered months were May and July (Fig. 4).

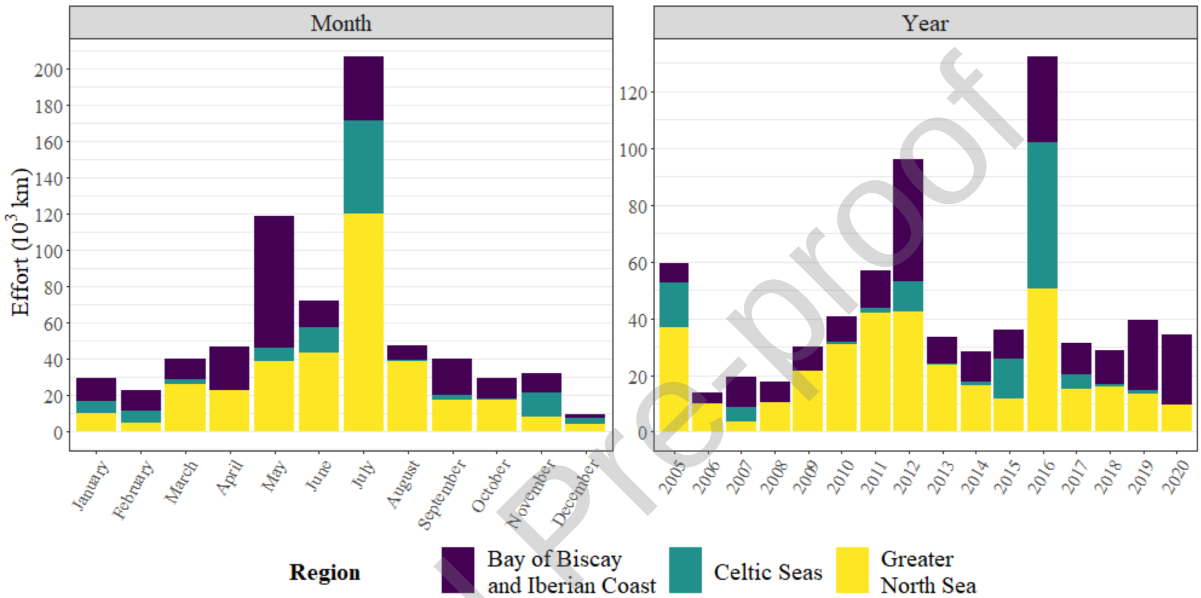


Figure 4. Cumulative monthly and annual survey effort from 2005 to 2020, separated by MSFD Region.

The surveyed areas were larger in June and July (Fig. 5). Throughout the year, surveys were more frequent in the shallow waters of the eastern/southern Greater North Sea and along the coast of the Bay of Biscay. Deep waters were mainly covered in January, June, July and September, and the coverage was lower than in shallow waters. The Portuguese offshore waters were not covered by the surveys.

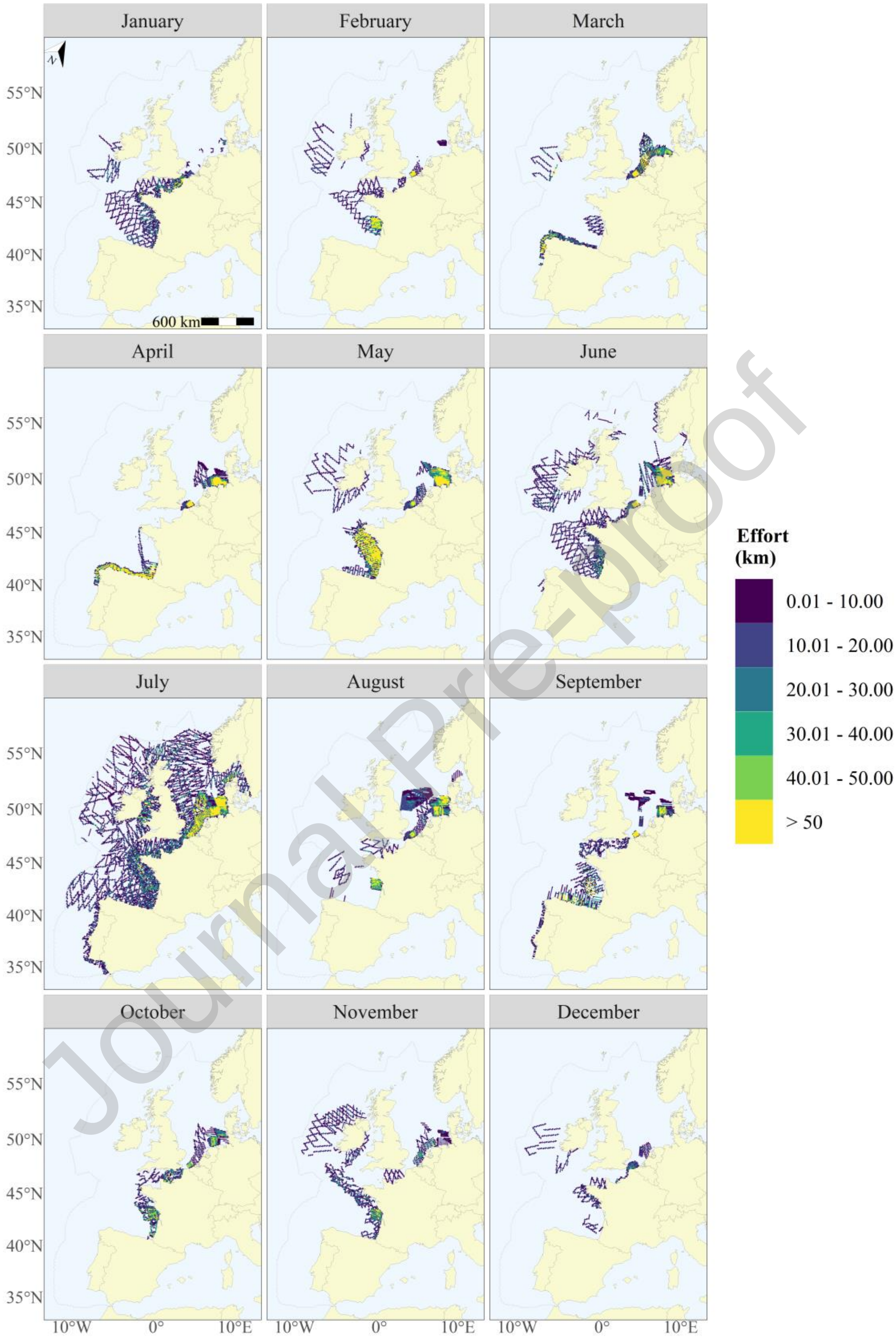


Figure 5. Geographic coverage of the surveys. Cumulative monthly effort per 10x10 km grid cell from 2005 to 2020.

3.2.Gap analysis in the environmental space

Monthly extrapolations and nearby data for 2019 are presented in the following sections, and the overall results are available at https://pelabox.univ-lr.fr/pelagis/Extrapolation/. Similar spatial and temporal patterns have been found over the years.

3.2.1 Extrapolations

Off the southern Iberian coast, predictions in deep waters were consistently extrapolated (Fig. 6). Predictions in the deep waters of the Bay of Biscay and Celtic Seas were extrapolated between January and May, although in some winter months (e.g., 2012, 2016, December 2019), they were partly interpolated. These deep waters were largely interpolated in summer.

The predictions were interpolated over the years in shallow waters of the Bay of Biscay, Iberian coast, the southern Greater North Sea and the southern Celtic Seas. In the north of the study area, from October to December-January, the interpolations gradually turned into extrapolations. These predictions reverted to interpolations from January to April: therefore, the number of extrapolations per month increased from [6,029: 14,067] in summer to [34,238: 54,929] in winter (minimum and maximum between 2005 and 2020), representing [8.1%: 18.9%] and [46.0%: 73.8%], respectively.

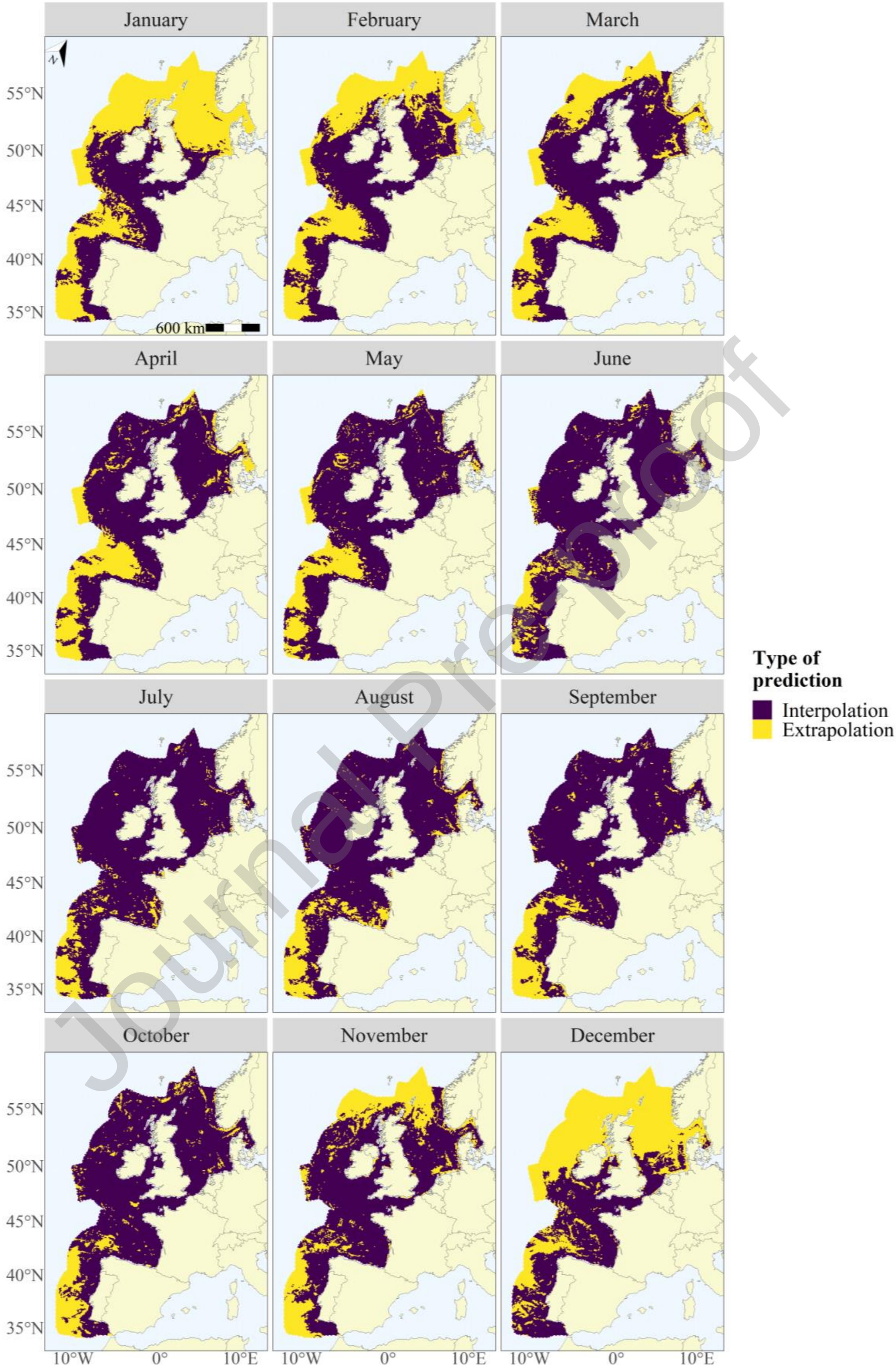


Figure 6. Extrapolation maps for 2019 based on survey data collected from 2005 to 2020 and environmental covariates in Table 1.

3.2.2. Nearby data

The percentages of nearby data in shallow waters (continental shelves) differed significantly from continental slopes, seamounts and abyssal plains (Fig. 7): the maximum percentages were 47.6% (35,381) in shallow waters, 2.0% (1,480) in abyssal plains, 1.4% (1,066) in continental slopes and 1.0% (744) in seamounts off the Iberian coast.

During the year, the nearby data have slowly shifted southward, from summer to winter. The highest percentages of nearby data decreased by 5% to 9% (3721 to 6699 nearby data) in January to April in shallow waters. They remained stable in abyssal plains, decreased to a minimum of 0.6% (447) on continental slopes from September to February, and to 0.7% (521) on seamounts (without seasonal pattern). The peak of nearby data was reached in summer in the Greater North Sea and Celtic Seas, and late spring/fall in the Bay of Biscay.

Variations in nearby data along the Iberian coast should be noted, with higher values from December to April than from July to September. Furthermore, some rare predictions totalled zero nearby data on the slope of the Portuguese coast from June to December. Some predictions also had zero nearby data in other areas: Portuguese offshore waters from December to April, the northern continental slope, and the entrance to Skagerrak (numbers increasing from November to February and decreasing until June).

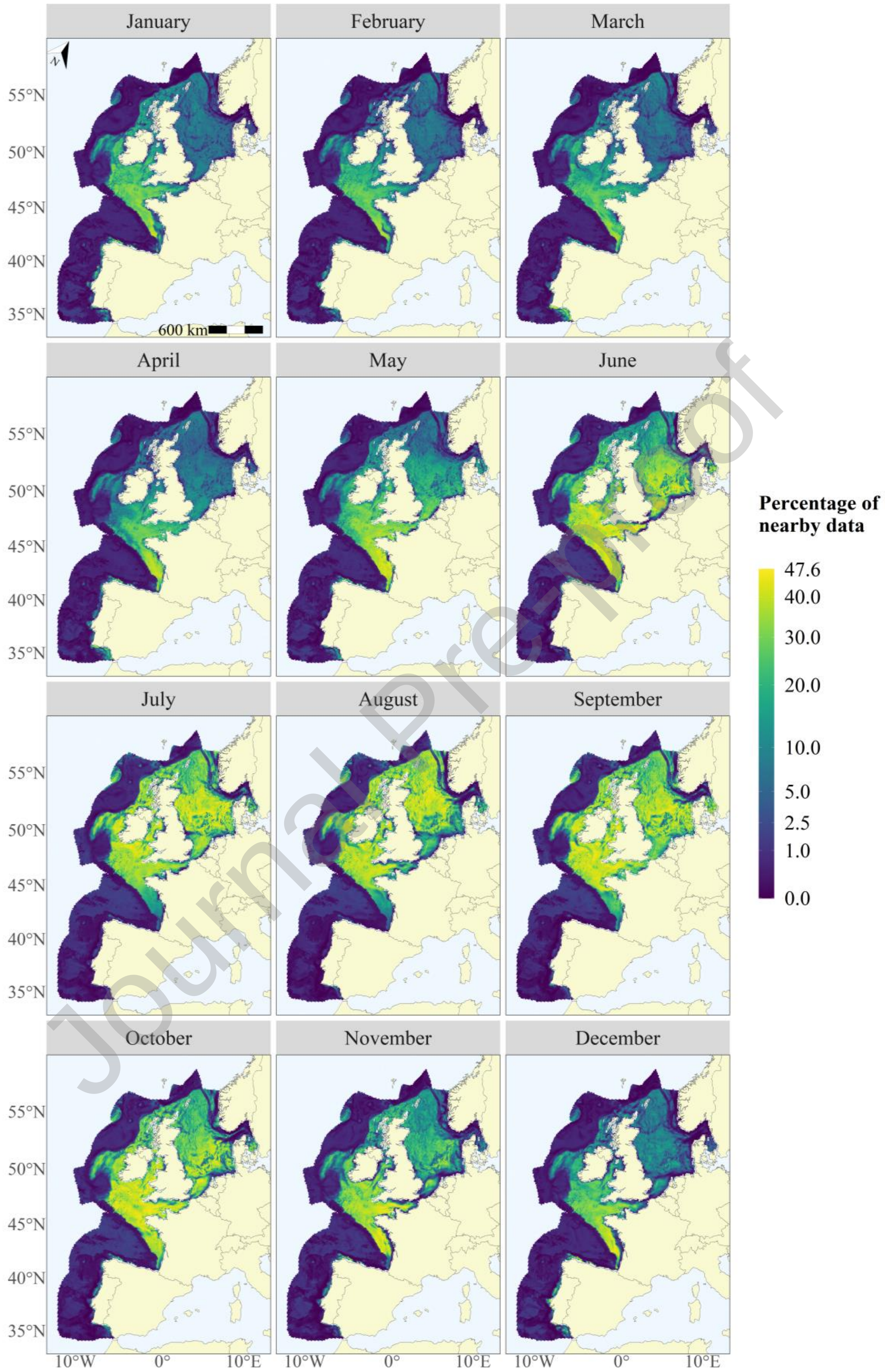


Figure 7. Percentages of nearby data of predictions for 2019 in the environmental space, based on survey data collected from 2005 to 2020 and environmental covariates in Table 1.

4. Discussion

4.1. Practical application: predictive reliability for cetacean abundances in the North-East Atlantic

This study highlights differences in the potential of the calibration data to build reliable distribution models over the study area, due to environmental gaps related to bathymetry, seabed slope, temperature, and primary productivity in the calibration dataset. Predictions in shallow waters without slope were particularly well supported by the calibration dataset. From May to November in the Bay of Biscay, the Celtic Seas, the English Channel and the North Sea, the regular and extensive survey coverage over months and years of the continental shelf has allowed to capture the annual variability of environments. These predictions are especially more data-driven and robust to modelling choices. Surveys were conducted less frequently and less extensively in winter and predictions during this season are based largely on data collected in northern regions during warmer months. Winter predictions of cetacean distribution are generally less reported due to uncertainties related to the potential differences between seasonal and regional species-environment relationships (Geelhoed et al., 2022; Gilles et al., 2016; Virgili et al., 2019). The nearby data show that predictions can be supported by a substantial part of the calibration dataset if the data can effectively be transferred between regions and seasons.

Environmental gaps in the calibration dataset were revealed in primary productivity drops and peaks, which were not fully captured by the surveys. This is the case in the northern study area, where surveys were only conducted in the summer and did not cover the cold and low productive waters from December to February, as well as the peaks of primary productivity in cold temperatures from April to May. This has led to a decrease in nearby data percentages in northern areas during this period and an increase in the number of extrapolations. High primary productivities in warm waters, found along the Iberian coast in summer, were also not given much support by this calibration dataset. These predictions were better supported in winter, when the primary productivity and temperature (but also the eddy kinetic energy) are lower. These environments are probably similar to those in which effort is greatest, i.e. shallow, temperate and less productive waters. The ecological rationale for transferring data from northern areas in warmer months to the Iberian coast should also be verified before predicting winter cetacean distribution in this area. Furthermore, the continental slope along the Iberian Peninsula limits the support of the calibration dataset to these predictions, as slopes represent a small area compared to continental shelves and abyssal plains in the North-East Atlantic and, apart from on the Iberian coast, are offshore and difficult to access. The temperature, primary productivity and slope, associated with the annual variability of the environment, led to some predictions having zero nearby data on the Portuguese coast, although surveys were conducted in July 2005, July 2016 and September 2019 in this area. This region is particularly important given its species richness and diversity (Correia et al., 2021; García-Barón et al., 2019) and would benefit from more surveys.

Overall, a large environmental gap was found on slopes in the calibration dataset, although the Bay of Biscay continental slope was, for example, regularly covered over the years and for several seasons (Authier et al., 2018; García-Barón et al., 2019; Hammond et al., 2013; 2021; Laran et al., 2017). In this area, many predictions were nonetheless interpolated throughout the year. However, the number of extrapolations on the northern slopes of the Celtic Seas and the North Sea increased sharply between summer and winter with the transition to cold and less productive environments.

Deep waters were also a source of significant environmental gaps, as percentages of nearby data were low, due to logistic constraints that limit the survey coverage of offshore areas. Large-scale, internationally coordinated surveys such as SCANS (1994, 2005, 2016 ; Hammond et al., 2002, 2013, 2021), CODA (CODA, 2009) and ObSERVE (Rogan et al., 2018) were substantially attributed to this coverage of deep waters and slopes. The effort was particularly higher in summer due to weather conditions. This is reflected in the large number of interpolations found in deep waters during the warm months. Nonetheless, the winter coverage effort (e.g. Laran et al., 2017; Rogan et al., 2018) has resulted in the interpolation of predictions in the covered areas and months. The large number of extrapolations in the Bay of Biscay abyssal plain during spring shows, however, that the surveys did not capture the cold, productive waters found in this area from February to May. Furthermore, the lack of surveys off Portugal has left a large environmental gap in this area during most of the year. Species-environment relationships in deep waters off Portugal may differ from those in the Bay of Biscay due to the presence of prominent topographical structures, such as seamounts (Cascão et al., 2020; Rovere et al., 2016). Predictions on seamount sides were the least supported by the calibration dataset due to the combination of slope, deep waters and the lack of surveys off Portugal where these environments occur. Their nearby data were data collected on continental slopes of the northern Iberian coast, Bay of Biscay, and Celtic seas. These predictions should be handled with caution and any management based on these predictions should caveat accordingly.

Since the distribution of cetaceans is strongly influenced by physiographic and oceanographic characteristics, related to the aggregation of prey and cetacean ecological restrictions (Cañadas et al., 2002; Kiszka et al., 2007; MacLeod et al., 2005; Virgili et al., 2019), environmental gaps related to temperature, primary productivity, bathymetry and seabed slope are highlighted, limiting the reliability of the predictions. Future surveys in geographical areas and periods representing these environments, especially offshore areas and in the winter and spring seasons, will contribute to fill these gaps and decrease the uncertainties. Surveys regularly conducted over the years have enabled to better capture the annual variability of environments and avoid extrapolations in years with unusual environmental conditions. These coverages are necessary to better understand cetacean distribution, in particular seasonally and in offshore areas, and allow effective monitoring of populations in the face of global change and ubiquitous human activities (Avila et al., 2018; Halpern et al., 2008, 2015). Variations in nearby data and extrapolations highlighted, in this analysis, the differences in seasonal survey coverage rather than monthly coverage, with cold seasons significantly less covered than warm seasons.

4.2.Considerations for SDMs predictive reliability

Our study did not reveal a significant influence of the eddy kinetic energy, gradient of temperature and aspect on the predictive reliability and robustness, although this does not mean that they have no influence. However, their variability at small rather than large scales, in contrast to bathymetry, primary productivity and temperature, as well as their lower seasonal variations could explain the consistency of the coverage over their range of values in different environments. Variations in reliability and robustness may occur due to these covariates but on a smaller scale and, therefore, not highlighted here. The nearby data (Fig. 7) show that the Gower's distance assigned a predominant weight in the calculation of nearby data to bathymetry due to its wide range, which attenuated differences in other covariates, such as temperature and primary productivity. The presence of high and uncommon primary productivity values in the dataset has further attenuated the differences between seasons and regions for this covariate, and preliminary data analysis may be required to rescale differences within the range of a covariate and between covariates.

Some choices may also influence this so-called gap analysis. Increasing the number of covariates, in the same way as using dynamic covariates (Mannocci et al., 2018), or a finer spatial or temporal resolution (Randin et al., 2009; Yates et al., 2018), increases the number of possible combinations of covariates, therefore increasing the degree of extrapolations (Authier et al., 2017). Model-users could also save the time required to compute extrapolations, nearby data and predictions by limiting the number of covariates and the spatio-temporal resolution *a priori*, depending on the modelling aims. The covariates making large contributions to extrapolations can be identified through existing tools (Bouchet et al., 2020) and removed during covariate selection prior to model fitting. Modifying the calibration dataset or the covariate set, however, will significantly change these results by increasing the percentages of nearby data and the number of interpolations when covariates are removed or calibration data added (Authier et al., 2017; Mannocci et al., 2018). Furthermore, the Gower's distance is relative to the ranges of covariates in the calibration set. Therefore, the percentages of nearby data might be low even in the case of small spatio-temporal extent, where environments are similar.

The ability to assess predictive reliability from the environmental similarity between the calibration data and the predictions is also expected to decrease with significant climate change, since the latter can lead to geographical shifts in environmental space (i.e. new range or average values for covariates in a given geographical area), open up future environments that are currently unsuitable for species or may result in species-environment relationship changes (Elith and Leathwick, 2009; Fitzpatrick and Hargrove, 2009; Veloz et al., 2012). Furthermore, species distribution may be determined by non-environmental factors, such as life cycles and residency patterns (Nathan et al., 2008). Hence, models often include covariates related to the geographical space, and modeller may consider including them as additional dimensions in the environmental space for the calculation of extrapolations and nearby data. This approach allows both geographical and environmental distances to be taken into account when analysing gaps in the calibration dataset used to fit distribution models. Ignorance maps, which consider temporal decay and geographical distance between calibration data and predictions (Rocchini et al., 2011; Tessarolo et al., 2021), may also provide valuable and relevant assessments in the aforementioned cases, although additional knowledge/assumptions are needed (e.g. decay range).

5. Conclusions

SDMs are invaluable tools for supporting decision-making for species conservation. Predictive uncertainties and errors due to dependence on modelling choices must therefore be appropriately communicated to improve their confidence in management, especially in the face of global change where new environments open up and restrain predictive reliability. We assessed the uncertainties of predictions from 2005 and 2020 for the first time in the combined regions of the Bay of Biscay, the North Sea, the Iberian Coast and the Celtic Seas for the common indicator *Abundance and distribution of cetaceans* of the OSPAR 2023 Quality Status Report (Geelhoed et al., 2022). Stakeholders can easily identify the predictions less informed by data and more by assumptions embedded in modelling choices with the extrapolations and nearby data (e.g., Figs. 6-7). These environmental gaps can be filled by future surveys conducted in the region and periods concerned, which would reduce the uncertainties in predicting cetacean distributions.

These metrics are useful tools for presenting caveating maps to end-users: they provide a descriptive summary of data available for SDM predictions, regardless of the modelling procedure. We recommend routinely reporting these metrics as additional information on predictive reliability before model-fitting (Bouchet et al., 2019), in complement to measures of predictive precision (e.g. coefficient of variations or standard error) that are available after model fitting.

CRediT authorship contribution statement

Rémi Pigeault: Formal analysis, Data curation, Software, Writing - original draft, Writing - review & editing, Visualization. **Matthieu Authier**: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing - review & editing, Visualization, Supervision, Funding acquisition. **Nadya C. Ramirez-Martinez**: Formal analysis, Data curation, Writing - review & editing, Methodology. **Auriane Virgili**: Formal analysis, Data curation, Software, Writing - review & editing, Methodology. **Steve C. V. Geelhoed**: Methodology, Data curation, Writing - review & editing. **Jan Haelters**: Methodology, Data curation, Writing - review & editing. **Maite Louzao**: Methodology, Data curation, Writing - review & editing. **Camilo Saaveedra**: Methodology, Data curation, Writing - review & editing. **Anita Gilles**: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing - review & editing, Visualization, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The results of extrapolations and nearby data, as well as the maps of the covariates used, are available on the shiny application: https://pelabox.univ-lr.fr/pelagis/Extrapolation/. The calibration dataset is available on the GitHub: https://github.com/osparcomm/Abundance-and-Distribution-of-Cetaceans (Geelhoed et al., 2022).

Acknowledgements

The authors thank the OSPAR Marine Mammal Expert Group (OMMEG) for valuable discussions and feedback, Chris Moulton from the OSPAR secretariat for assistance with the data call, the French Ministry in Charge of Ecology for financial support and Mathieu Genu for the design of the Shiny application. SCVG thanks Rijkswaterstaat for the financial support of his OSPAR QSR work. AG thanks the German Federal Agency for Nature Conservation for funding her time as the lead of OMMEG (grant no. 3521532207). ML was partially supported by a Ramón y Cajal contract (RYC-2012-09897) from the Spanish Government, as well as the PETFISH project, funded by the European Maritime, Fisheries and Aquaculture Fund (EMFAF) under the Data Collection Framework Regulation. RP and NCRM were partially supported by project CoastalFutures, funded by the German Federal Ministry of Education and Research (BMBF) under grant number 03F0911H. The responsibility for the content of this publication lies with the authors.

We further acknowledge the contribution of the ObSERVE data originating from the ObSERVE Project undertaken by the Department of Communications, Climate Action and Environment and the National Parks and Wildlife Service, Department of Culture, Heritage and the Gaeltacht in Ireland. We thank Philip S. Hammond for providing the SCANS data (i.a., SCANS-II LIFE04NAT/GB/000245) and all observers, institutions and staff responsible for these programmes and their respective funders. The national surveys in Germany were funded by the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection, the Federal Agency for Nature Conservation and the Federal Ministry of Food and Agriculture. The Dutch surveys were funded by the Dutch Ministry of Economic Affairs, Ministry of Agriculture, Nature and Food, the Ministry of Infrastructure and the Environment (Rijkswaterstaat). We thank Signe Sveegaard for providing data from the Danish monitoring of porpoises and the Danish Environmental Protection Agency for funding these surveys. Belgian survey data result from the monitoring programme on the ecological effects of offshore wind farms in Belgian waters. The national surveys in France were funded by the Ministry in charge of Ecology and Office Français de la Biodiversité, with support from Ifremer for the ship surveys. For the surveys in Spain, we thank all crew and scientists participating in BIOMAN and JUVENA surveys on R/Vs ‘Ramon Margalef′ and ‘Emma Bardan’. This study has been partially funded by the EMFF (European Maritime Fishing Funds) through the Data collection framework, the “Departamento de Desarrollo Económico, Sostenibilidad y Medio Ambiente-Viceconsejería de Agricultura, Pesca y Política Alimentaria, Dirección de Pesca y Acuicultura” of the Basque Government and the Spanish General Secretary for Fisheries. PELACUS and IBERAS were partially funded by the EMFF and the Ministry of Ecological Transition and Demographic Challenge through a contract with the IEO for the implementation of monitoring programs to respond to the Marine Strategy Framework Directive. We thank two anonymous reviewers for their constructive comments on an earlier draft. This Open Access publication was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 491094227 "Open Access Publication Funding" and the University of Veterinary Medicine Hannover, Foundation.

This study has been conducted using E.U. Copernicus Marine Service Information; https://doi.org/10.48670/moi-00021, https://doi.org/10.48670/moi-00019.

References

Authier, M., Dorémus, G., Van Canneyt, O., Boubert, J.-J., Gautier, G., Doray, M., Duhamel, E., Massé, J., Petitgas, P., Ridoux, V., Spitz, J. (2018). Exploring change in the relative abundance of marine megafauna in the Bay of Biscay, 2004–2016. *Progress in Oceanography 166*, 159–167. https://doi.org/10.1016/j.pocean.2017.09.014

Authier, M., Saraux, C., Péron, C. (2017). Variable selection and accurate predictions in habitat modelling: a shrinkage approach. *Ecography 40*, 549–560. https://doi.org/10.1111/ecog.01633

Avila, I. C., Kaschner, K., & Dormann, C. F. (2018). Current global risks to marine mammals: taking stock of the threats. *Biological Conservation*, 221, 44-58. https://doi.org/10.1016/j.biocon.2018.02.021

Becker, E.A., Carretta, J.V., Forney, K.A., Barlow, J., Brodie, S., Hoopes, R., Jacox, M.G., Maxwell, S.M., Redfern, J.V., Sisson, N.B., Welch, H., Hazen, E.L. (2020). Performance evaluation of cetacean species distribution models developed using generalized additive models and boosted regression trees. *Ecol Evol 10*, 5759–5784. https://doi.org/10.1002/ece3.6316

Bouchet, P.J., Miller, D.L., Roberts, J.J., Mannocci, L., Harris, C.M., Thomas, L. (2020). dsmextra: Extrapolation assessment tools for density surface models. *Methods Ecol. Evol. 11*, 1464–1469. https://doi.org/10.1111/2041-210X.13469

Bouchet, P.J., Miller, D.L., Roberts, J.J., Mannocci, L., Harris, C.M., Thomas, L. (2019). From here and now to there and then: Practical recommendations for extrapolating cetacean density surface models to novel conditions: 59.

Buckland, S.T. (Ed.) (2009). Introduction to distance sampling: estimating abundance of biological populations, Reprinted. ed. *Oxford Univ. Press*, Oxford.

Cañadas, A., Sagarminaga, R., García-Tiscar, S. (2002). Cetacean distribution related with depth and slope in the Mediterranean waters off southern Spain. *Deep Sea Research Part I: Oceanographic Research Papers 49*, 2053–2073. https://doi.org/10.1016/S0967-0637(02)00123-1

Cascão, I., Lammers, M. O., Prieto, R., Santos, R. S., & Silva, M. A. (2020). Temporal patterns in acoustic presence and foraging activity of oceanic dolphins at seamounts in the Azores. *Scientific reports*, *10*(1), 3610. https://doi.org/10.1038/s41598-020-60441-4

Correia AM, Sousa-Guedes D, Gil Á, Valente R, Rosso M, Sousa-Pinto I, Sillero N and Pierce GJ (2021). Predicting Cetacean Distributions in the Eastern North Atlantic to Support Marine Management. *Front. Mar. Sci. 8*:643569. doi: 10.3389/fmars.2021.643569

Dormann, C.F. (2011). Modelling Species' Distributions, in: Jopp, F., Reuter, H., Breckling, B. (Eds.), Modelling Complex Ecological Dynamics. *Springer Berlin Heidelberg*, Berlin, Heidelberg, pp. 179–196. https://doi.org/10.1007/978-3-642-05029-9_13

Elith, J., Leathwick, J.R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annu. Rev. Ecol. Evol. Syst. 40*, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Fitzpatrick, M.C., Hargrove, W.W. (2009). The projection of species distribution models and the problem of non-analog climate. *Biodivers Conserv 18*, 2255–2261. https://doi.org/10.1007/s10531-009-9584-8

García-Barón, I., Authier, M., Caballero, A., Vázquez, J.A., Santos, M.B., Murcia, J.L., Louzao, M. (2019). Modelling the spatial abundance of a migratory predator: A call for transboundary marine protected areas. *Divers Distrib 25*, 346–360. https://doi.org/10.1111/ddi.12877

Geelhoed, S., Authier, M., Pigeault, R., Gilles, A. (2022). *Abundance and Distribution of Cetaceans. I*n: OSPAR, 2023: The 2023 Quality Status Report for the Northeast Atlantic. OSPAR Commission, London. Available at: https://oap.ospar.org/en/ospar-assessments/quality-status-reports/qsr-2023/indicator-assessments/abundance-distribution-cetaceans/

Gilles, A., Viquerat, S., Becker, E.A., Forney, K.A., Geelhoed, S.C.V., Haelters, J., Nabe-Nielsen, J., Scheidat, M., Siebert, U., Sveegaard, S., Beest, F.M., Bemmelen, R., Aarts, G. (2016). Seasonal habitat-based density models for a marine top predator, the harbor porpoise, in a dynamic environment. *Ecosphere 7*. https://doi.org/10.1002/ecs2.1367

Guisan, A., Tingley, R., Baumgartner, J.B., Naujokaitis-Lewis, I., Sutcliffe, P.R., Tulloch, A.I.T., Regan, T.J., Brotons, L., McDonald-Madden, E., Mantyka-Pringle, C., Martin, T.G., Rhodes, J.R., Maggini, R., Setterfield, S.A., Elith, J., Schwartz, M.W., Wintle, B.A., Broennimann, O., Austin, M., Ferrier, S., Kearney, M.R., Possingham, H.P., Buckley, Y.M. (2013). Predicting species distributions for conservation decisions. *Ecol Lett 16*, 1424–1435. https://doi.org/10.1111/ele.12189

Halpern, B.S., Frazier, M., Potapenko, J., Casey, K.S., Koenig, K., Longo, C., Lowndes, J.S., Rockwood, R.C., Selig, E.R., Selkoe, K.A., Walbridge, S. (2015). Spatial and temporal changes in cumulative human impacts on the world's ocean. *Nat Commun 6*, 7615. https://doi.org/10.1038/ncomms8615

Halpern, B.S., Walbridge, S., Selkoe, K.A., Kappel, C.V., Micheli, F., D'Agrosa, C., Bruno, J.F., Casey, K.S., Ebert, C., Fox, H.E., Fujita, R., Heinemann, D., Lenihan, H.S., Madin, E.M.P., Perry, M.T., Selig, E.R., Spalding, M., Steneck, R., Watson, R. (2008). A Global Map of Human Impact on Marine Ecosystems. *Science 319*, 948–952. https://doi.org/10.1126/science.1149345

Hammond, P., Lacey, C., Gilles, A., Viquerat, S., Börjesson, P., Herr, H., Macleod, K., Ridoux, V., Santos, M., Teilmann, J., Vingada, J., Øien, N. (2021). Estimates of cetacean abundance in European Atlantic waters in summer 2016 from the SCANS-III aerial and shipboard surveys 42.

CODA (2009). *Cetacean Offshore Distribution and Abundance in the European Atlantic (CODA)*. Final report to the European Commission. St Andrews: Sea Mammal Research Unit. Available at: https://www.npws.ie/sites/default/files/publications/pdf/Anon_2009_CODA_Final_Report.pdf

Hammond, P.S., Berggren, P., Benke, H., Borchers, D.L., Collet, A., Heide-J, M.P., Heimlich, S., Hiby, A.R., Leopold, M.F., Øien, N. (2002). Abundance of harbour Porpoise and other cetaceans in the North Sea and adjacent waters 16.

Hammond, P.S., Macleod, K., Berggren, P., Borchers, D.L., Burt, L., Cañadas, A., Desportes, G., Donovan, G.P., Gilles, A., Gillespie, D., Gordon, J., Hiby, L., Kuklik, I., Leaper, R., Lehnert, K., Leopold, M., Lovell, P., Øien, N., Paxton, C.G.M., Ridoux, V., Rogan, E., Samarra, F., Scheidat, M., Sequeira, M., Siebert, U., Skov, H., Swift, R., Tasker, M.L., Teilmann, J., Van Canneyt, O., Vázquez, J.A. (2013). Cetacean abundance and distribution in European Atlantic shelf waters to inform conservation and management. *Biological Conservation 164*, 107–122. https://doi.org/10.1016/j.biocon.2013.04.010

Hutchinson, G. E. (1957). "Concluding remarks. *Cold Spring Harb Symp Quant Biol*, 415-427.

King, G., Zeng, L. (2007). When Can History Be Our Guide? The Pitfalls of Counterfactual Inference. *Int Studies Q 51*, 183–210. https://doi.org/10.1111/j.1468-2478.2007.00445.x

Kiszka, J., Heithaus, M., Wirsing, A. (2015). Behavioural drivers of the ecological roles and importance of marine mammals. *Mar. Ecol. Prog. Ser. 523*, 267–281. https://doi.org/10.3354/meps11180

Kiszka, J., Macleod, K., Van Canneyt, O., Walker, D., Ridoux, V. (2007). Distribution, encounter rates, and habitat characteristics of toothed cetaceans in the Bay of Biscay and adjacent waters from platform-of-opportunity data. *ICES Journal of Marine Science 64*, 1033–1043. https://doi.org/10.1093/icesjms/fsm067

Lambert, C., Pettex, E., Dorémus, G., Laran, S., Stéphan, E., Canneyt, O.V., Ridoux, V. (2017). How does ocean seasonality drive habitat preferences of highly mobile top predators? Part II: The eastern North-Atlantic. *Deep Sea Research Part II: Topical Studies in Oceanography 141*, 133–154. https://doi.org/10.1016/j.dsr2.2016.06.011

Laran, S., Authier, M., Blanck, A., Doremus, G., Falchetto, H., Monestiez, P., Pettex, E., Stephan, E., Van Canneyt, O., Ridoux, V. (2017). Seasonal distribution and abundance of cetaceans within French waters- Part II: The Bay of Biscay and the English Channel. *Deep Sea Research Part II: Topical Studies in Oceanography 141*, 31–40. https://doi.org/10.1016/j.dsr2.2016.12.012

MacLeod, C.D., Bannon, S.M., Pierce, G.J., Schweder, C., Learmonth, J.A., Herman, J.S., Reid, R.J. (2005). Climate change and the cetacean community of north-west Scotland. *Biological Conservation 124*, 477–483. https://doi.org/10.1016/j.biocon.2005.02.004

Mannocci, L., Roberts, J.J., Halpin, P.N., Authier, M., Boisseau, O., Bradai, M.N., Cañadas, A., Chicote, C., David, L., Di-Méglio, N., Fortuna, C.M., Frantzis, A., Gazo, M., Genov, T., Hammond, P.S., Holcer, D., Kaschner, K., Kerem, D., Lauriano, G., Lewis, T., Notarbartolo di Sciara, G., Panigada, S., Raga, J.A., Scheinin, A., Ridoux, V., Vella, A., Vella, J. (2018). Assessing cetacean surveys throughout the Mediterranean Sea: a gap analysis in environmental space. *Sci Rep 8*, 3126. https://doi.org/10.1038/s41598-018-19842-9

Mesgaran, M.B., Cousens, R.D., Webber, B.L. (2014). Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Diversity Distrib. 20*, 1147–1159. https://doi.org/10.1111/ddi.12209

Moore, S.E. (2008). Marine mammals as ecosystem sentinels. *Journal of Mammalogy 89, 534–540*. https://doi.org/10.1644/07-MAMM-S-312R1.1

Nathan, R., Getz, W. M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D., & Smouse, P. E. (2008). A movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences, 105*(49), 19052–19059. https://doi.org/10.1073/pnas.0800375105

Owens, H.L., Campbell, L.P., Dornak, L.L., Saupe, E.E., Barve, N., Soberón, J., Ingenloff, K., Lira-Noriega, A., Hensz, C.M., Myers, C.E., Peterson, A.T. (2013). Constraints on interpretation of ecological niche models by limited environmental ranges on calibration areas. *Ecological Modelling 263*, 10–18. https://doi.org/10.1016/j.ecolmodel.2013.04.011

Palacios, D. M., S. J. Bograd, D. G. Foley, and F. B. Schwing (2006). Oceanographic characteristics of biological hot spots in the North Pacific: a remote sensing perspective. *Deep- Sea Research II: Topical Studies in Oceanography* 53:250–269. https://doi.org/10.1016/j.dsr2.2006.03.004

Palacios, D. M., M. F. Baumgartner, K. L. Laidre, and E. J. Gregr. (2013). Beyond correlation: integrating environmentally and behaviourally mediated processes in models of marine mammal distributions. *Endangered Species Research 22*:191–203. https://doi.org/10.3354/esr00558

Qiao, H., Feng, X., Escobar, L.E., Peterson, A.T., Soberón, J., Zhu, G., Papeş, M. (2019). An evaluation of transferability of ecological niche models. *Ecography 42*, 521–534. https://doi.org/10.1111/ecog.03986

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <URL: https://www.R-project.org/>

Randin, C.F., Engler, R., Normand, S., Zappa, M., Zimmermann, N.E., Pearman, P.B., Vittoz, P., Thuiller, W., Guisan, A. (2009). Climate change and plant distribution: local models predict high-elevation persistence. *Global Change Biology 15, 1557–1569*. https://doi.org/10.1111/j.1365-2486.2008.01766.x

Redfern, J.V., Ferguson, M., Becker, E.A., Hyrenbach, K.D., Good, C., Barlow, J., Kaschner, K., Baumgartner (2006). *Marine Ecology Progress Series 310:271*. Mar Ecol Prog Ser 25.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jiménez-Valverde, A., Ricotta, C., Bacaro, G., Chiarucci, A. (2011). Accounting for uncertainty when mapping species distributions: The need for maps of ignorance. *Progress in Physical Geography: Earth and Environment 35*, 211–226. https://doi.org/10.1177/0309133311399491

Rogan, E., Breen, P., Mackey, M., Cañadas, A., Scheidat, M., Geelhoed, S. & Jessopp, M. (2018). Aerial surveys of cetaceans and seabirds in Irish waters: Occurrence, distribution and abundance in 2015-2017. Department of Communications, Climate Action & Environment and National Parks and Wildlife Service (NPWS), Department of Culture, Heritage and the Gaeltacht, Dublin, Ireland. 297pp.

Rogan, E., Cañadas, A., Macleod, K., Santos, M.B., Mikkelsen, B., Uriarte, A., Van Canneyt, O., Vázquez, J.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. https://doi.org/10.1016/j.dsr2.2017.03.015

Rousseau, J.S., Betts, M.G. (2022). Factors influencing transferability in species distribution models. *Ecography 2022*. https://doi.org/10.1111/ecog.06060

Rovere, M., Bo, M., Alessi, J., Paoli, C., Villani, N., Vassallo, P., ... & Roccatagliata, N. (2016). Seamounts and seamount-like structures of the Tyrrhenian Sea. Atlas of the Mediterranean seamounts and seamount-like structures. Publisher: IUCN, Gland, Switzerland and Málaga, Spain. Editors: Maurizio Würtz, Marzia Rovere, 111-184.

Scheidat, M., Gilles, A., Kock, K., Siebert, U. (2008). Harbour porpoise Phocoena phocoena abundance in the southwestern Baltic Sea. *Endang. Species Res. 5*, 215–223. https://doi.org/10.3354/esr00161

Taheri, S., Naimi, B., Rahbek, C., Araújo, M.B. (2021). Improvements in reports of species redistribution under climate change are required. *Sci. Adv. 7*, eabe1110. https://doi.org/10.1126/sciadv.abe1110

Tessarolo, G., Ladle, R.J., Lobo, J.M., Rangel, T.F., Hortal, J. (2021). Using maps of biogeographical ignorance to reveal the uncertainty in distributional data hidden in species distribution models. *Ecography 44*, 1743–1755. https://doi.org/10.1111/ecog.05793

Veloz, S.D., Williams, J.W., Blois, J.L., He, F., Otto-Bliesner, B., Liu, Z. (2012). No-analog climates and shifting realized niches during the late quaternary: implications for 21st-century predictions by species distribution models. *Glob Change Biol 18*, 1698–1713. https://doi.org/10.1111/j.1365-2486.2011.02635.x

Virgili, A., Authier, M., Boisseau, O., Cañadas, A., Claridge, D., Cole, T., Corkeron, P., Dorémus, G., David, L., Di-Méglio, N., Dunn, C., Dunn, T.E., García-Barón, I., Laran, S., Lauriano, G., Lewis, M., Louzao, M., Mannocci, L., Martínez-Cedeira, J., Palka, D., Panigada, S., Pettex, E., Roberts, J.J., Ruiz, L., Saavedra, C., Santos, M.B., Van Canneyt, O., Vázquez Bonales, J.A., Monestiez, P., Ridoux, V. (2019).

Combining multiple visual surveys to model the habitat of deep-diving cetaceans at the basin scale: Large-scale modelling of deep-diving cetacean habitats. *Global Ecol Biogeogr 28*, 300–314. https://doi.org/10.1111/geb.12850

Virgili, A., Racine, M., Authier, M., Monestiez, P., Ridoux, V. (2017). Comparison of habitat models for scarcely detected species. *Ecological Modelling 346*, 88–98. https://doi.org/10.1016/j.ecolmodel.2016.12.013

Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S., Fielding, A.H., Bamford, A.J., Ban, S., Barbosa, A.M., Dormann, C.F., Elith, J., Embling, C.B., Ervin, G.N., Fisher, R., Gould, S., Graf, R.F., Gregr, E.J., Halpin, P.N., Heikkinen, R.K., Heinänen, S., Jones, A.R., Krishnakumar, P.K., Lauria, V., Lozano-Montes, H., Mannocci, L., Mellin, C., Mesgaran, M.B., Moreno-Amat, E., Mormede, S., Novaczek, E., Oppel, S., Ortuño Crespo, G., Peterson, A.T., Rapacciuolo, G., Roberts, J.J., Ross, R.E., Scales, K.L., Schoeman, D., Snelgrove, P., Sundblad, G., Thuiller, W., Torres, L.G., Verbruggen, H., Wang, L., Wenger, S., Whittingham, M.J., Zharikov, Y., Zurell, D., Sequeira, A.M.M. (2018). Outstanding Challenges in the Transferability of Ecological Models. *Trends in Ecology & Evolution 33*, 790–802. https://doi.org/10.1016/j.tree.2018.08.001

Zurell, D., Elith, J., Schröder, B. (2012). Predicting to new environments: tools for visualizing model behaviour and impacts on mapped distributions: Predicting to new environments. *Diversity and Distributions 18*, 628–634. https://doi.org/10.1111/j.1472-4642.2012.00887.x

Zurell, D., Franklin, J., König, C., Bouchet, P.J., Dormann, C.F., Elith, J., Fandos, G., Feng, X., Guillera-Arroita, G., Guisan, A., Lahoz-Monfort, J.J., Leitão, P.J., Park, D.S., Peterson, A.T., Rapacciuolo, G., Schmatz, D.R., Schröder, B., Serra-Diaz, J.M., Thuiller, W., Yates, K.L., Zimmermann, N.E., Merow, C. (2020). A standard protocol for reporting species distribution models. *Ecography 43*, 1261–1277. https://doi.org/10.1111/ecog.04960

Figures and tables

Table 1. Candidate environmental covariates used to model the cetacean distributions. Source A: EMODnet Bathymetry 2020 Digital Terrain Model (https://www.emodnet-bathymetry.eu/). Slope and aspect were derived from bathymetry using the *Terrain* function (Raster R-package). Source B: Copernicus (https://resources.marine.copernicus.eu): Global Ocean Physics Reanalysis (B_1), Global Ocean Biogeochemistry Hindcast (B_2; see Acknowledgements for DOIs). The SST gradients were calculated from the SST means, using the *DetectFronts* function (*Grec* R package).

Environmental variable		Original spatial resolution	Original temporal resolution	Temporal prediction resolution	Abbreviation	Source	Justification
Physiographic	Bathymetry (m)	1/16 arc minute	NA	NA	Bathy	A	Proxy for cetacean prey distribution (e.g., the deep-divers feed on squid and fish in the deep-water column).
	Slope (rad)				Slope		Associated with currents, high slopes induce enhanced primary production or prey aggregation.
	Aspect (rad)				Aspect		Describe currents and prominent structures such as canyons, seamounts or mountain chains, used as proxies for predator hotspots and useful in locations where access to biological data is limited.
Oceanographic	Sea surface temperature mean (°C) and gradient (°C/m)	0.083° (-5500 m to 0 m)	Monthly	Monthly	mSST, gradSST	B_1	Variability over time and horizontal gradients of SST reveal front locations, mixing of water and is associated with enhanced primary production and prey aggregations.
	Eddy kinetic Energy (m/s)				EKE		High EKE relates to the development of eddies, upwelling of nutrients and enhanced primary production, which induce prey aggregation.

	Net primary productivity (mg.m^{-3}. day^{-1})	0.25° (-5500 m to 0 m)			NPPV	B_2	Net primary production is a proxy of zooplankton distribution, feeding cetacean prey.

Figure 8. Representation of the environmental space covered by an illustrative dataset with two environmental covariates, temperature and bathymetry. Extrapolations (prediction *q*) are predictions under environments outside the environmental coverage (red line) in the environmental space, and interpolations (prediction *p*) are predictions inside the environmental coverage. The more nearby data a prediction has in the environmental space, the more this prediction is based on empirical data and is reliable. This figure results from a simulation using Gower's distance in the *WhatIf* R package (King and Zeng, 2007; version 1.5-8). Unlike Euclidean distances, the radius around a Gower's distance produces a diamond shape (distances in each dimension are not squared before being added together; see Methods).

Figure 9. Flowchart of data preparation steps and analyses. The dashed arrows represent additional steps in the prediction of abundances.

Figure 10. Study area delimited by OSPAR regions II, III and IV and the overlapping MSFD sub-regions (EU Marine Strategy Framework Directive). Bathymetry source: EMODnet DTM (https://www.emodnet-bathymetry.eu/).

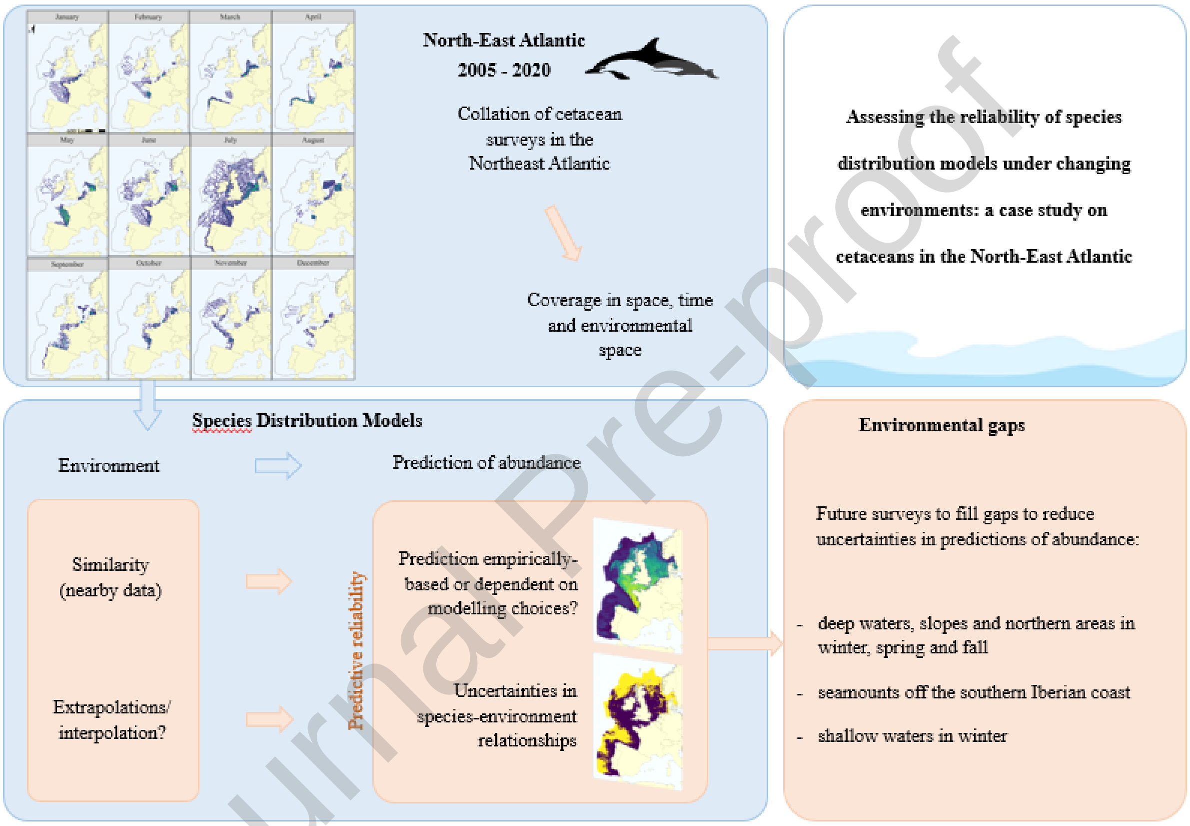
Figure 11. Cumulative monthly and annual survey effort from 2005 to 2020, separated by MSFD Region.

Figure 12. Geographic coverage of the surveys. Cumulative monthly effort per 10x10 km grid cell from 2005 to 2020.

Figure 13. Extrapolation maps for 2019 based on survey data collected from 2005 to 2020 and environmental covariates in Table 1.

Figure 14. Percentages of nearby data of predictions for 2019 in the environmental space, based on survey data collected from 2005 to 2020 and environmental covariates in Table 1.

Graphical abstract



Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

- **Ethics Statement**

  x Not applicable: This manuscript does not include human or animal research.

  ☐If this manuscript involves research on animals or humans, it is imperative to disclose all approval details.

Highlights

- 2005-2020 cetacean survey data collected in the North-East Atlantic were aggregated
- We study environments where predictions depend on model assumptions
- The data support was large in shallow waters but decreased sharply in deeper waters
- Variations in temperature, primary productivity and slope influenced the data support
- This deepens our understanding of the potential to build reliable distribution models