Environmental variables for distribution modelling of UK marine megafauna species

First published October 2022

Natural England Commissioned Report NECR447

www.gov.uk/natural-england

Natural England Commissioned Report NECR447

Environmental variables for distribution modelling of UK marine megafauna species

Darren Wilkinson & John Quinn

School of Biological, Earth and Environmental Sciences, University College Cork

Published October 2022

Project details

Natural England commission a range of reports from external contractors to provide evidence and advice to assist us in delivering our duties. The views in this report are those of the authors and do not necessarily represent those of Natural England.

Natural England Project manager

Alex Banks

Contractor

University College Cork, Ireland

Authors

Darren Wilkinson[1] and Professor John L. Quinn[1,2]

[1]School of Biological, Earth and Environmental Sciences, University College Cork.

[2]Environmental Research Institute, University College Cork, Cork, Ireland

Keywords

UK marine megafauna, species distribution models, conservation protocols

Citation

This report should be cited as:

WILKINSON, D., and QUINN, J. L. 2022. *Environmental variables for distribution modelling of UK marine megafauna species.* NECR447. Natural England.

Further information

This report can be downloaded from the Natural England Access to Evidence Catalogue: **http://publications.naturalengland.org.uk/**. For information on Natural England publications contact the Natural England Enquiry Service on 0300 060 3900 or e-mail **enquiries@naturalengland.org.uk**.

Executive summary

The demand for predictive distribution models for marine species has grown dramatically in recent years for the purposes of conservation and marine spatial planning. The aim of this report is to summarise the environmental predictor variables regularly used in distribution models for marine megafauna species around the UK. It also identifies the variables that are most frequently retained in the final models, and which ones are removed during model selection. This report summarises the results for different marine taxa (seabirds, cetaceans, seals, basking sharks, and turtles).

In addition to providing a breakdown of the environmental variables used in species distribution models, this report outlines the modelling techniques typically used.

Some degree of caution should be taken when interpreting the analysis in this report as the ability to detect the effects of environmental variables is likely influenced by the modelling approaches taken and by the amount of species occurrence data available. Nevertheless, this report highlights a wide range of variables that can be targeted in future modelling studies of marine megafauna species with some degree of confidence based on the extent of their successful use to date.

Contents

1. Introduction

Understanding the distribution of species is crucial for the effective planning and implementation of conservation protocols. There is a growing demand for predictive distribution models for marine species due to the continuing threats faced by marine biodiversity such as pollution (Thushari & Senevirathna, 2020), invasive species (Dias et al., 2019), fisheries bycatch (Žydelis et al., 2009), climate change (Hoegh-Guldberg & Bruno, 2010), and others. Where extensive distribution data is lacking for a species, distribution models can provide important insight into the species' habitat preferences and potential conflict with anthropogenic threats. Species distribution models (SDMs) are empirical models that aim to predict the distribution of a species by combining data on its presence or abundance with environmental variables (Elith & Leathwick, 2009). The use of SDMs for marine species was relatively rare until the early 2000s (Elith & Leathwick, 2009), largely due to the difficulties posed with studying marine ecosystems.

One major challenge is the three-dimensional nature of the marine environment (Bentlage et al., 2013), which adds complexity to the modelling process. Typically, a combination of static (e.g. depth, seabed slope) and dynamic (e.g. sea surface temperature, chlorophyll *a* concentration) environmental variables are used in SDMs. These dynamic variables are usually derived from remote-sensing processes (e.g. Breen et al., 2017; McClellan et al., 2014; Rogan et al., 2017), although most only represent the upper layers of the water column (Melo-Merino et al., 2020). Therefore, SDM accuracy is impeded for pelagic species that are found in mid-water habitats (Bentlage et al., 2013). In addition, the dynamism of the marine ecosystem presents a challenge as it can result in spatial or temporal lags between the environmental state and the species' response (Redfern et al., 2006). Consequently, the environmental predictors used in SDMs must be carefully selected to ensure they reflect the spatio-temporal scales in which animal-environment interactions take place (Scales et al., 2017). Also, in order to construct SDMs, data on the presence or abundance of the study species is required but detection of highly mobile marine megafauna is often difficult (Elith & Leathwick, 2009). This can be due to factors such as the weather conditions during boat or land-based surveys, the fact that some species travel vast distances, and that others spend prolonged periods of time below the surface of the water (Redfern et al., 2006). Studies using distribution modelling techniques in the marine ecosystem have been biased towards coastal and shallow waters where surveying is easier to conduct (Robinson et al., 2011).

Modelling Techniques

Many different modelling techniques have been developed to examine species distributions. The methods vary in terms of how they select environmental variables, measure the relative contribution of each variable, and the predictive power of the model (Elith et al., 2006). Correlative modelling is one of the common techniques employed. This method correlates measures of species occurrence (presence-only, presence-absence, or abundance) with environmental variables to predict distribution and habitat suitability (see review by Guisan & Zimmermann, 2000). Other modelling techniques used include 'hybrid' models, which combine correlative and process-based models (Smolik et al., 2010), and mechanistic procedures, which include functional traits of the study species (e.g. morphology and physiology) and environmental data (Kearney & Porter, 2009). Mechanistic modelling, while providing a mechanistic understanding of underlying processes that are not explained through correlative methods (Kearney & Porter, 2009), are not frequently used in the marine ecosystem as they require large amounts of data (Elith & Leathwick, 2009). Instead, correlative models are most commonly featured in SDM studies as they are relatively simple to construct and don't require much data (Robinson et al., 2011), which is a major advantage when researching understudied marine taxa. In a review conducted by Melo-Merino et al. (2020) on SDMs in marine environments, it was found that correlative techniques were used in 307 of the 328 studies featured. Marine species represented in these studies ranged from seagrass and planktonic organisms to megafauna species such as cetaceans and seabirds (Melo-Merino et al., 2020).

While outside the scope of this study, it is important to note that SDMs conducted for cetaceans using observation data should include a detectability function to account for perception bias of species occurrence. This can arise due to marine mammal behaviour and to observer bias caused by factors such as weather conditions. Despite being a common feature of cetacean SDM studies, detectability functions are not typically performed for distribution modelling of other marine taxa such as seabirds.

Correlative Approaches

Regression-based models are one of the most commonly used statistical techniques for modelling species distribution. There is an extensive range of regression techniques that vary in their assumptions of parameter distribution and the functional form of the relationships between variables; however, all methods work by modelling variation in measures of species occurrence (presence-only or presence-absence) or abundance (count data) with one or more environmental variables. Linear regression is the simplest form and the models produced by this method tend to be relatively easy to interpret and apply. Generalised Linear Models (GLMs; Nelder & Wedderburn, 1972) have been used for a long period of time to examine relationships between species occurrence/abundance and habitat variables. This type of model uses a link function to produce a linear relationship between the response and predictor variables. GLMs show great flexibility in how they handle different types of response variables, for example binary data (e.g. presence-absence) can be modelled using logistic regression, while count data can be modelled with a Poisson regression. GLMs assume that the relationship between the response and predictor variable is parametric, although this may not be the case for all relationships between species and their habitat. Generalised Additive Models (GAMs; Hastie & Tibshirani, 1986) are non-parametric extensions of GLMs that use a smoothing function instead of a linear function. Using GAMs to produce SDMs provides additional flexibility for the fitting of non-linear relationships that in many cases are more ecologically realistic than linearity; however, overfitting can be an issue with GAMs.

The focus of many SDM studies has shifted to predicting species distribution and a number of methods have been developed especially for predictions. These include the machine-learning methods of maximum entropy (maxent; Phillips et al., 2006), genetic algorithms (Stockwell & Peters, 1999), and classification and regression trees.

Regardless of the technique employed, model selection is a crucial step in the modelling process. Early SDMs used statistical tests based on p-values to determine which explanatory variables to retain in the model. However, in more recent times, methods such as Akaike's information criterion (AIC) and Bayesian information criterion (BIC) have been developed. These model selection criteria evaluate several candidate models to determine the combination of variables that provides the best fit. In addition, interpretation of SDMs is easier if the variables are not correlated. Multicollinearity occurs when environmental variables are correlated and is often managed by removing one of the correlated variables from the model and retaining the explanatory variable that is hypothesised or known to be more ecologically relevant to the study species.

This report provides an overview of the environmental predictor variables commonly used in SDM studies of marine megafauna species around the United Kingdom (UK) and indicates which variables should be considered for future modelling. Caution should be taken when interpreting the analysis in this report as the ability to detect effects is likely influenced by the modelling approaches taken and by the amount of species occurrence data available.

2. Methodology

This report is based on the analysis of published results from research articles. Potential studies for inclusion in this analysis were found by searching all databases in Clarivate's Web of Science using the following search criteria:

Topic = (seabird* OR "marine mammal*" OR cetacean* OR pinniped* OR elasmobranch* OR turtle* OR shark* OR whale* OR dolphin* OR porpoise* OR seal*) AND (distribution OR abundance OR track* OR at-sea observation) AND (environmental variable* OR environmental predictor* OR spatial model*) AND ("Celtic Sea" OR "English Channel" OR "Irish Sea" OR " North Sea" OR "North Atlantic").

Timespan = all years.

Refined by: Subject Areas = (Environmental Sciences Ecology OR Zoology OR Marine Freshwater Biology OR Biodiversity Conservation OR Behavioral Sciences OR Oceanography OR Fisheries OR Evolutionary Biology).

The search resulted in 1393 studies being identified for inclusion in the analysis. Despite the detailed search criteria, many of the papers found featured species not relevant to this report or the sampling occurred in water bodies outside of the area of interest. The 1393 papers were subsequently manually examined and only studies that modelled the distribution of marine megafauna species using at least one environmental variable in water bodies surrounding the United Kingdom (Celtic Sea, English Channel, Irish Sea, North Sea, and North-East Atlantic Ocean) were retained for analysis. This yielded 44 papers containing 144 distribution models (Appendix 1). Information on the study species, type of modelling technique used, variables included in the initial model, and variables retained in the final model following model selection, were extracted. Variables that were included in SDMs as offset terms or correction factors (usually a measure of survey effort) were not included in the analysis. Variables that were removed due to multicollinearity with another variable were considered to be part of the initial model, but not retained in the final distribution model. While interaction terms were not common in the studies featured in this analysis, where they occurred, the individual variables comprising the interaction were considered to have been included in the model.

3. Results

All the models represented in this study were constructed using correlative methods (Table 1). Statistical modelling techniques were most common with GAMs (including Generalised Additive Mixed Models, GAMMs) accounting for 44% of the models represented in this study. GLMs (including Generalised Linear Mixed Models, GLMMs) accounted for a further 15%. Maxent was the most frequently used machine-learning method, with 37 of the 144 models using this method.

Table 1. The number of times and the percentage of the total number of models each modelling technique was featured in the selected studies. GAM = Generalised Additive Model; GAMM = Generalised Additive Mixed Model; Maxent = Maximum Entropy; GLM = Generalised Linear Model; GLMM = Generalised Linear Mixed Model; GEE = Generalised Estimating Equations; ENFA = Ecological Niche Factor Analysis; EENM = Ensemble Ecological Niche Model (the model indicated below integrated GLM, Multiple Adaptive Regression Splines, and Generalised Boosting Model approaches); GARP = Genetic Algorithm for Rule Set Production; PCA = Principal Component Analysis

Model Type	No. of models	% of Total
GAM (including GAMM)	63	43.75
Maxent	37	25.69
GLM (including GLMM)	22	15.28
Classification Tree	9	6.25
GEE	7	4.86
ENFA	2	1.39
EENM	1	0.69
GARP	1	0.69
PCA	1	0.69
Spearman’s Rank Correlation	1	0.69

In total, 64 different predictor variables were included in the 144 models, and these were categorised into 8 groups (Table 2). The static predictor variables of depth (n=129) and seabed slope (n=94) were featured most often in the initial models prior to selection (Table 2). Depth was retained in 89 models (69%) and slope in 56 (60%). Sea surface temperature was the most used dynamic variable (Table 2), initially occurring in 90 models with a retention rate of 68%. Geographic coordinates, distance to bathymetric contours, and salinity, despite featuring in many models (47, 40, and 43 respectively), were retained less than 50% of the time (Table 2).

Table 2. The number of times each environmental predictor variable was used in the initial and final (following model selection) SDMs and the retention rate expressed as a percentage. The environmental predictor variables have been divided into 8 groups, and within each group the variables are sorted according to the number of initial models they featured in.

Variable	No. of initial models	No. of final models	Retained %
Atmospheric Variables			
Wind Speed	2	2	100
Wind Direction	1	0	0
Sea Level Pressure	1	0	0
Biochemical/Chemical Variables			
Chlorophyll *a* Concentration	64	43	67.2
Salinity	43	21	48.8
Surface Fluorescence	1	0	0
Ecological Variables			
Prey Abundance/Distribution	48	28	58.3
Primary Productivity	4	0	0
Presence of Other Species	1	1	100
Geographic Variables			
Distance to Coast	70	50	71.4
Geographic Coordinates	47	22	46.8
Distance to Bathymetric Contour	40	18	45

Variable	No. of initial models	No. of final models	Retained %
Distance to Colony/Nest	25	24	96
Distance to Haul-out Site	7	7	100
Survey Site	4	4	100
Distance to Oceanographic Front	4	3	75
Distance to Intertidal Zone	3	3	100
Distance to Estuary	1	1	100
Distance to Prey	1	1	100
Seabed Topographic Variables			
Depth	129	89	69
Slope	94	56	59.6
Seabed Sediment	30	22	73.3
Rugosity	18	8	44.4
Aspect	13	7	53.8
Seabed Hardness	6	6	100
Hydrodynamic Variables			
Sea Surface Temperature	90	61	67.8
Current Speed	25	15	60
Mixed Layer Depth	17	10	58.8
Spring-Neap Tide Cycle	13	8	61.5
Sea Surface Height	12	8	66.7
Tidal State	10	3	30
Tidal Power	8	5	62.5

Variable	No. of initial models	No. of final models	Retained %
Sea State	7	5	71.4
Turbulence	6	6	100
Presence/Frequency of Oceanographic Front	6	4	66.7
Change in Tide Height	4	3	75
Oceanographic Front Gradient Density	4	3	75
Current Direction	4	1	25
Tidal Stratification	3	2	66.7
Spring Tidal Amplitude	3	0	0
Current Level	2	0	0
Swell	2	0	0
Current Magnitude	1	1	100
Oceanographic Front Persistence	1	1	100
Tide Height	1	1	100
Water Clarity	1	1	100
Water Mass	1	1	100
Side of Oceanographic Front	1	0	0
Temporal Variables			
Year	21	14	66.7
Day	15	11	73.3
Month	8	7	87.5
Time of Day	8	5	62.5
Hour	7	4	57.1

Variable	No. of initial models	No. of final models	Retained %
Season	2	2	100
Day Length	1	1	100
Time to High Tide	1	1	100
Other Variables			
Anthropogenic Noise/Activity	9	5	55.6
Survey Effort	3	3	100
Survey Method	3	2	66.7
Observer ID	2	1	50
Backscattering Strength	1	1	100
Glare	1	1	100
Observer Visibility	1	1	100
Vertical Shear	1	1	100

Seabirds

Forty-five of the models represented seabird species and these models contained 28 different predictor variables (Table 3). Depth was the most frequently used variable (n=39) and was retained in 25 models. Prey abundance/distribution was the next most common, occurring in 32 initial models and, following model selection, was retained in 21 (66%). Geographic variables featured frequently in the seabird models. Distance to the colony and coast were found to be extremely important variables for predicting seabird distribution. Both featured in 25 models with distance to the colony being retained 96% of the time, and distance to the coast was present in 22 final models (88%). In contrast, some geographic variables were not found to be useful predictors of seabird distribution. Distance to bathymetric contours (e.g. 200 metre isobath) was used in 8 models but was not retained in any. Geographic coordinates were only retained in 20% of the 25 models it was featured in.

The biochemical/chemical variables of chlorophyll *a* concentration and salinity were identified as influential environmental predictors of seabird distribution. In addition, while not occurring regularly in the initial models, the variables sea surface height (n=7), current speed (n=6), seabed hardness (n=6) and turbulence (n=6) were selected in the final models 100% of the time.

Table 3. The number of times each environmental predictor variable was used in the initial and final (following model selection) seabird SDMs and the retention rate expressed as a percentage.

Variable	No. of initial models	No. of final models	Retained %
Depth	39	25	64.1
Prey Abundance/Distribution	32	21	65.6
Sea Surface Temperature	31	20	64.5
Distance to Colony/Nest	25	24	96
Distance to Coast	25	22	88
Geographic Coordinates	25	5	20
Slope	11	10	90
Chlorophyll *a* Concentration	11	9	81.8
Salinity	9	8	88.9
Rugosity	8	6	75
Distance to Bathymetric Contour	8	0	0
Sea Surface Height	7	7	100
Current Speed	6	6	100
Seabed Hardness	6	6	100
Turbulence	6	6	100
Seabed Sediment	4	2	50
Distance to Intertidal Zone	3	3	100
Tidal Power	3	1	33.3
Tidal State	3	0	0
Day	2	2	100

Variable	No. of initial models	No. of final models	Retained %
Hour	2	2	100
Aspect	1	1	100
Month	1	1	100
Presence/Frequency of Oceanographic Front	1	1	100
Water Clarity	1	1	100
Water Mass	1	1	100
Year	1	1	100
Primary Productivity	1	0	0

Dolphins and Porpoises

Dolphins and porpoises featured in 55 of the 144 models included in this study. Forty-five predictor variables were used (Table 4). Topographic variables were repeatedly used in the models exploring dolphin/porpoise distribution (Table 4). Depth (n=50) and slope (n=49) were the two most commonly used predictors. Following model selection, depth was included in 66% of the final models, but slope only appeared in 49%. Seabed sediment (n=12) and aspect (n=10) also regularly occurred in the initial models and had a retention rate of 67% and 60%, respectively. Out of the variables that featured in 10 or more models, geographic coordinates had the highest retention rate (79%), and salinity had the lowest (30%).

Table 4. The number of times each environmental predictor variable was used in the initial and final (following model selection) dolphin and porpoise SDMs and the retention rate expressed as a percentage.

Variable	No. of initial models	No. of final models	Retained %
Depth	50	33	66
Slope	49	24	49
Sea Surface Temperature	35	22	62.9
Distance to Coast	30	20	66.7
Chlorophyll *a* concentration	29	17	58.6

Variable	No. of initial models	No. of final models	Retained %
Salinity	23	7	30.4
Distance to Bathymetric Contour	20	11	55
Geographic Coordinates	14	11	78.6
Mixed Layer Depth	12	8	66.7
Seabed Sediment	12	8	66.7
Year	11	6	54.5
Day	10	7	70
Aspect	10	6	60
Current Speed	9	6	66.7
Prey Abundance/Distribution	8	0	0
Spring-Neap Tide Cycle	7	4	57.1
Time of Day	7	4	57.1
Sea State	6	5	83.3
Anthropogenic Noise/Activity	6	4	66.7
Tidal State	6	2	33.3
Survey Site	4	4	100
Current Direction	4	1	25
Hour	4	1	25
Distance to Oceanographic Front	3	2	66.7
Oceanographic Front Gradient Density	3	2	66.7
Tidal Stratification	3	2	66.7
Spring Tide Amplitude	3	0	0

Variable	No. of initial models	No. of final models	Retained %
Month	2	2	100
Season	2	2	100
Presence/Frequency of Oceanographic Front	2	1	50
Survey Method	2	1	50
Tidal Power	2	1	50
Current Level	2	0	0
Change in Tide Height	1	1	100
Day Length	1	1	100
Distance to Estuary	1	1	100
Distance to Prey	1	1	100
Presence of Other Species	1	1	100
Tide Height	1	1	100
Wind Speed	1	1	100
Sea Level Pressure	1	0	0
Side of Oceanographic Front	1	0	0
Surface Fluorescence	1	0	0
Swell	1	0	0
Wind Direction	1	0	0

Whales

Whale species featured in 24 models and 33 predictor variables were used (Table 5). There were several similarities to the dolphin/porpoise results. Once again, depth (n=23) and slope (n=22) were the most used variables with retention rates of 78% and 55%, respectively. Geographic coordinates also appeared to be important as it was kept in 75% of the 8 models in which its inclusion was examined.

Sea surface temperature was utilised in 15 models and was retained in the final model 11 times. Chlorophyll *a* concentration was also found to be important, occurring in 10 final models.

Table 5. The number of times each environmental predictor variable was used in the initial and final (following model selection) whale SDMs and the retention rate expressed as a percentage.

Variable	No. of initial models	No. of final models	Retained %
Depth	23	18	78.3
Slope	22	12	54.5
Sea Surface Temperature	15	11	73.3
Chlorophyll *a* concentration	15	10	66.7
Geographic Coordinates	8	6	75
Year	7	5	71.4
Distance to Bathymetric Contour	7	4	57.1
Distance to Coast	7	2	28.6
Current Speed	7	1	14.3
Rugosity	10	2	20
Spring-Neap Tide Cycle	6	4	66.7
Month	5	4	80
Seabed Sediment	5	4	80
Mixed Layer Depth	5	2	40
Salinity	5	1	20
Sea Surface Height	4	1	25
Survey Effort	3	3	100
Change in Tide Height	3	2	66.7
Prey Abundance/Distribution	3	2	66.7

Variable	No. of initial models	No. of final models	Retained %
Primary Productivity	3	0	0
Anthropogenic Noise/Activity	2	1	50
Day	2	1	50
Aspect	2	0	0
Glare	1	1	100
Hour	1	1	100
Presence/Frequency of Oceanographic Front	1	1	100
Tidal Power	1	1	100
Time of Day	1	1	100
Observer Visibility	1	1	100
Wind Speed	1	1	100
Observer ID	1	0	0
Sea State	1	0	0
Swell	1	0	0

Seals

Twelve of the 144 models focused on seal distribution. Twenty predictor variables featured, but only 6 were used more than once (Table 6). Depth (n=10), seabed sediment (n=8), distance to haul-out site (n=7), and slope (n=6) all appear to be important variables for the prediction of seal distribution.

Table 6. The number of times each environmental predictor variable was used in the initial and final (following model selection) seal SDMs and the retention rate expressed as a percentage.

Variable	No. of initial models	No. of final models	Retained %
Depth	10	8	80
Seabed Sediment	8	7	87.5

Variable	No. of initial models	No. of final models	Retained %
Distance to Haul-out Site	7	7	100
Slope	6	6	100
Current Speed	3	2	66.7
Distance to Coast	3	2	66.7
Backscattering Strength	1	1	100
Current Magnitude	1	1	100
Day	1	1	100
Observer ID	1	1	100
Salinity	1	1	100
Tidal Power	1	1	100
Tidal State	1	1	100
Time to High Tide	1	1	100
Vertical Shear	1	1	100
Year	1	1	100
Anthropogenic Noise/Activity	1	0	0
Chlorophyll *a* Concentration	1	0	0
Sea Surface Height	1	0	0
Sea Surface Temperature	1	0	0

Basking Sharks

Basking sharks were the subject of 7 SDMs. Sixteen predictor variables were used (Table 7), and despite the small quantity of models, it appears that some of these variables are important predictors of basking shark distribution. Sea surface temperature was retained in all 7 final models and chlorophyll *a* concentration was present in 6 final models. Depth, and to a less extent seabed slope, seem to influence basking shark distribution. In addition, prey abundance/distribution, distance to bathymetric contours, distance to the coast, salinity, and measures of oceanographic front activity show evidence of being influential predictor variables.

Table 7. The number of times each environmental predictor variable was used in the initial and final (following model selection) basking shark SDMs and the retention rate expressed as a percentage.

Variable	No. of initial models	No. of final models	Retained %
Sea Surface Temperature	7	7	100
Chlorophyll *a* Concentration	7	6	85.7
Depth	6	5	83.3
Slope	5	3	60
Prey Abundance/Distribution	4	4	100
Distance to Bathymetric Contour	4	3	75
Distance to Coast	4	3	75
Salinity	4	3	75
Presence/Frequency of Oceanographic Front	2	1	50
Distance to Oceanographic Front	1	1	100
Oceanographic Front Gradient Density	1	1	100
Oceanographic Front Persistence	1	1	100
Seabed Sediment	1	1	100
Survey Method	1	1	100
Tidal Power	1	1	100
Year	1	1	100

Marine Turtles

Only one SDM was constructed for a marine turtle species. Eight variables were included in the initial model but only 6 were retained in the final version – chlorophyll *a* concentration, distance to the coast, prey abundance/distribution, salinity, sea surface temperature, and slope (Table 8).

Table 8. The number of times each environmental predictor variable was used in the initial and final (following model selection) marine turtle SDMs and the retention rate expressed as a percentage.

Variable	No. of initial models	No. of final models	Retained %
Chlorophyll *a* Concentration	1	1	100
Distance to Coast	1	1	100
Prey Abundance/Distribution	1	1	100
Salinity	1	1	100
Sea Surface Temperature	1	1	100
Slope	1	1	100
Depth	1	0	0
Distance to Bathymetric Contour	1	0	0

4. Discussion

This report provides an overview of the commonly used environmental predictor variables in SDM studies of marine megafauna species around the UK and an indication of which variables should be considered for different marine taxa. Interpretation of this overview should be done with caution because the ability to detect effects is likely influenced by the modelling approaches taken and by the amount of species occurrence data available.

Modelling methods and occurrence data

The correlative techniques of statistical and machine-learning modelling dominate the literature on species distribution modelling of marine megafauna around the UK. None of the studies featured in this report employed the alternative hybrid (Smolik *et al*., 2010) or mechanistic (Kearney & Porter, 2009) SDM methods. A major limitation on the distribution modelling of many species in the marine environment is the lack of data on their occurrence and insufficient knowledge of their ecology (Bentlage *et al*., 2013). This has restricted the type of modelling that has been conducted and the species for which SDMs could be generated. In recent times, data collection on the movements and behaviour of marine megafauna has greatly improved through the enhancement and refinement of existing approaches, and the development of new methods (e.g. Nowacek *et al*., 2016). Technological advances have also increased the range of species for which accurate occurrence data can be gathered. For example, Global Positioning System (GPS) devices that weigh less than 1g are now available and can be deployed on the smallest seabird species breeding in the UK, the ~28g European Storm-petrel *Hydrobates pelagicus* (Bolton, 2020). As a result of new and improved data collection methods, there are now large data repositories containing extensive data on species occurrence (e.g. Movebank, MegaMove). Improved species occurrence data will enhance modelling procedures and will allow researchers to construct better predictive distribution models for marine species under future climate conditions or planned anthropogenic activity (e.g. offshore wind farm developments).

Environmental variables

An extensive list of environmental variables has been used to model the distribution of marine species. Melo-Merino *et al*. (2020) conducted a review on the use of SDMs on all marine taxa in the world's oceans and identified the inclusion of 173 different variables; Tremblay *et al*. (2009), focusing just on seabird species, reported the use of 101 environmental explanatory variables; and this study detected 64 different variables that have been used for the modelling of marine megafauna distributions around the UK. Environmental data have become readily accessible through global databases such as those provided by the National Oceanic and Atmospheric Administration (NOAA) and the General Bathymetric Chart of the Oceans (GEBCO). With all this data available, it can be tempting for modellers to construct SDMs by including many potential environmental predictor variables and trust that model selection will identify which ones influence the study species' distribution (Elith & Leathwick, 2009); however, the inclusion of too many environmental parameters in the model can result in overfitting (Peterson *et al*., 2007). Consequently, there is a strong argument for the inclusion of only environmental variables that are believed to be ecologically relevant to the study species (Elith & Leathwick, 2009).

Despite the vast amount of environmental data that is available for modelling studies, the environmental variables used in marine SDMs are typically restricted to sea surface variables (e.g. sea surface temperature, sea surface salinity) which are often derived from remote-sensing at high spatial resolutions or from modelled data, and measurements from the seabed (e.g. depth, seabed sediment type). Only a small proportion of the environmental data used in SDMs represent the mid-layers of the water column (e.g. mixed layer depth). The lack of environmental data within the water column can impact the accuracy of SDMs, especially for marine mammals that spend time both at the surface and underwater.

Marine megafauna species

Seabirds

Distance to the colony and the coast were found to be prominent predictors of seabird distribution. This is not surprising as the majority of seabird studies conducted in the UK are during the breeding season, as this is the only time in the year when many seabird species are present on land and so are accessible for monitoring and research. Breeding seabirds are central place foragers, and their distribution is constrained by the requirement to return to the nest at regular intervals to care for their chicks (Quillfeldt *et al*., 2010). During the breeding season, seabirds need to support the energetic demands of their offspring, while maintaining their own body condition (Burke & Montevecchi, 2009), and as a result, the distribution and abundance of prey can strongly influence seabird distribution. Prey resources for megafauna species in the marine environment are often associated with bathymetric features such as shelf edges (Cox *et al*., 2018) and this likely provides the ecological explanation for the importance of topographic variables such as slope and rugosity, and of chlorophyll *a* concentration (proxy for primary productivity), in the seabird SDMs. It is expected that outside the breeding season, the importance of some predictor variables will diminish, while others will increase in predictive power.

Cetaceans: whales, dolphins, and porpoises

Topographic variables appear to be key predictors of the distribution of cetaceans around the UK. The influence of topography is probably a result of its impact on prey distribution and concentration (Naud *et al*., 2013). Sea surface temperature was also identified as an important environmental variable for cetacean SDMs. This is likely due to the fact that sea surface temperature is known to influence the distribution of sandeels *Ammodytes tobinaus*, an important prey item of cetaceans, around the UK (van der Kooij *et al*., 2008). In contrast to other marine megafauna taxa, geographic coordinates were regularly retained in the final models of dolphin, porpoise, and whale distribution. Latitude and longitude can be included in SDMs to act as a proxy for other variables such as distance from the coast, water masses, and bathymetric regions. However, they sometimes act, intentionally or inadvertently, as proxies for unmeasured environmental variables (Redfern *et al*., 2006), and other unidentified factors influencing the distribution of cetacean species around the UK are likely. Despite these generalisations, the influence of environmental variables is highly likely to vary significantly among cetacean species due to factors such as body size and foraging strategies, and especially whether the species is oceanic or coastal.

Seals

A small number of models in this study featured grey and/or harbour seals (n=12). Like cetaceans, topographic variables had a notable influence on their distribution. Once again, this relationship is likely due to the impact topography has on the distribution of prey species. Distance to the haul-out site was retained in each of the models it was featured in. This affiliation was expected as hauling-out (temporarily moving onto land) is a common behaviour in pinnipeds and is performed for many reasons, including breeding and moulting. Outside of the breeding and moulting periods, grey seals still spend over 40% of their time on or near a haul-out site (McConnell *et al*., 1999). Despite only being included in one model, measures of tide (e.g. tidal power, tidal state, time to high tide) showed evidence of being good predictors of seal distribution and should be examined further in future analyses.

Basking sharks and turtles

Only a few SDM studies were found in the literature search for basking sharks (n=7) and turtle species (n=1). More distribution modelling is required to test which environmental variables are important for these marine megafauna species in UK waters. For basking sharks, the dynamic variables of sea surface temperature, chlorophyll *a* concentration, and salinity were found to influence the distribution during the summer months in this highly migratory species. As a planktivorous species, the relationship with primary productivity (i.e. chlorophyll *a* concentration) was expected. Variables of oceanographic front activity should be looked at in future modelling studies as the initial evidence suggests they may have a bearing on basking shark distribution.

5. Final remarks

This report summarises published research on the distribution of marine megafauna species in UK waters and has highlighted the range of predictor variables commonly used in species distribution models. The importance of variables is dependent on phylogeny, or more accurately, on the species' position in the food chain, their mode of foraging, and the type of prey consumed, the availability of which will vary considerably over space and time and at different scales. Effect size was not within the scope of this report and could have an impact on cost-benefit analyses of including variables in future monitoring. Nevertheless, as the demand grows for predictive modelling of marine species for conservation and marine spatial planning, this report highlights a wide range of variables that can be targeted for each marine megafauna taxa with some degree of confidence based on their utility to date.

List of tables

Appendices

Appendix 1. SDM studies included in the environmental variable analysis

Aarts, G., MacKenzie, M., McConnell, B., Fedak, M., Matthiopoulos, J., 2008. Estimating space-use and habitat preference from wildlife telemetry data. *Ecography*, **31**, 140-160.

Anderwald, P., Evans, P.G.H., Dyer, R., Dale A., Wright, P.J., Hoelzel, A.R., 2012. Spatial scale and environmental determinants in minke whale habitat use and foraging. *Marine Ecology Progress Series*, **450**, 259-274.

Arso Civil, M., Quick, N.J., Cheney, B., Pirotta, E., Thompson, P.M., Hammond, P.S., 2019. Changing distribution of the east coast of Scotland bottlenose dolphin population and the challenges of area-based management. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **29**, 178-196.

Austin, R.A., Hawkes, L.A., Doherty, P.D., Henderson, S.M., Inger, R., Johnson, L., Pikesley, S.K., Soldandt, J.-L., Speedie, C., Witt, M.J., 2019. Predicting habitat suitability for basking sharks (*Cetorhinus maximus*) in UK waters using ensemble ecological niche modelling. *Journal of Sea Research*, **153**, 101767.

Bailey, H., Hammond, P.S., Thompson, P.M., 2014. Modelling harbour seal habitat by combining data from multiple tracking systems. *Journal of Experimental Marine Biology and Ecology*, **450**, 30-39.

Bailey, H., Thompson, P.M., 2009. Using marine mammal habitat modelling to identify priority conservation zones within a marine protected area. *Marine Ecology Progress Series*, **378**, 279-287.

Baines, M., Reichelt, M., Griffin, D., 2017. An autumn aggregation of fin (*Balaenoptera physalus*) and blues whales (*B. musculus*) in the Porcupine Seabight, southwest of Ireland. *Deep-Sea Research Part II: Topical Studies in Oceanography*, **141**, 168-177.

Breen, P., Brown, S., Reid, D., Rogan, E., 2016. Modelling cetacean distribution and mapping overlap with fisheries in the northeast Atlantic. *Ocean & Coastal Management*, **134**, 140-149.

Breen, P., Brown, S., Reid, D., Rogan, E., 2017. Where is the risk? Integrating a spatial distribution model and a risk assessment to identify areas of cetacean interaction with fisheries in northeast Atlantic. *Ocean & Coastal Management*, **136**, 148-155.

Brookes, K.L., Bailey, H., Thompson, P.M., 2013. Predictions from harbor porpoise habitat association models are confirmed by long-term passive acoustic monitoring. *Journal of the Acoustical Society of America*, **14**, 2523-2533.

Cox, S.L., Witt, M.J., Embling, C.B., Godley, B.J., Hosegood, P.J., Miller, P.I., Votier, S.C., Ingram, S.N., 2017. Temporal patterns in habitat use by small cetaceans at an oceanographically dynamic renewable energy test site in the Celtic Sea. *Deep-Sea Research Part II: Topical Studies in Oceanography*, **141**, 178-190.

de Boer, M.N., Simmonds, M.P., Reijnders, P.J.H., Aarts, G., 2014. The influence of topographic and dynamic cyclic variables on the distribution of small cetaceans in a shallow coastal system. *PLoS ONE*, **9**, e86331.

Embling, C.B., Gillibrand, P.A., Gordon, J., Shrimpton, J., Stevick, P.T., Hammond, P.S., 2010. Using habitat models to identify suitable sites for marine protected areas for harbour porpoises (*Phocoena phocoena*). *Biological Conservation*, **143**, 267-279.

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., van Beest, F.M., van Bemmelen, R., Aarts, G., 2016. Seasonal habitat-based density models for a marine top predator, the harbor porpoise, in a dynamic environment. *Ecosphere*, **7**, e01367.

Gremillet, D., Gallien, F., El Ksabi, N., Courbin, N., 2020. Sentinels of coastal ecosystems: the spatial ecology of European shags breeding in Normandy. *Marine Biology*, **167**, 43.

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., Vásquez, J.A., 2013. Cetacean abundance and distribution in European Atlantic shelf waters to inform conservation and management. *Biological Conservation*, **164**, 107-122.

Hastie, G.D., Russell, D.J.F., Lepper, P., Elliott, J., Wilson, B., Benjamins, S., Thompson, D., 2017. Harbour seals avoid tidal turbine noise: Implications for collision risk. *Journal of Applied Ecology*, **55**, 684-693.

Hastie, G.D., Swift, R.J., Slesser, G., Thompson, P.M., Turrell, W.R., 2005. Environmental models for predicting oceanic dolphin habitat in the Northeast Atlantic. *ICES Journal of Marine Science*, **62**, 760-770.

Huon, M., Jones, E.L., Matthiopoulos, J., McConnell, B., Caurant, F., Vincent, C., 2015. Habitat selection of Gray Seals (*Halichoerus grypus*) in a Marine Protected Area in France. *The Journal of Wildlife Management*, **79**, 1091-1100.

Isojunno, S., Matthiopoulos, J., Evans, P.G.H., 2012. Harbour porpoise habitat preferences: robust spatio-temporal inferences from opportunistic data. *Marine Ecology Progress Series*, **448**, 155-170.

Johnston, A., Thaxter, C.B., Austin, G.E., Cook, A.S.C.P., Humphreys, E.M., Still, D.A., Mackey, A., Irvine, R., Webb, A., Burton, N.H.K., 2015. Modelling the abundance and distribution of marine birds accounting for uncertain species identification. *Journal of Applied Ecology*, **52**, 150-160.

Jones, E.L., Sparling, C.E., McConnell, B.J., Morris, C.D., Smout, S., 2017. Fine-scale harbour seal usage for informed marine spatial planning. *Scientific Reports*, **7**, 11581.

Kane, A., Pirotta, E., Wischnewski, S., Critchley, E.J., Bennison, A., Jessopp, M., Quinn, J.L., 2010. Spatio-temporal patterns of foraging behaviour in a wide-ranging seabird reveal the role of primary productivity in locating prey. *Marine Ecology Progress Series*, **646**, 175-188.

Lieber, L., Nimmo-Smith, W.A.M., Waggitt, J.J., Kregting, L., 2018. Fine-scale hydrodynamic metrics underlying predator occupancy patterns in tidal stream environments. *Ecological Indicators*, **94**, 397-408.

MacLeod, C.D., Mandleberg, L., Schweder, C., Bannon, S.M., Pierce, G.J., 2008a. A comparison of approaches for modelling the occurrence of marine animals. *Hydrobiologia*, **612**, 21-32.

MacLeod, C.D., Weir, C.R., Begoña Santos, M., Dunn, T.E., 2008b. Temperature-based summer habitat partitioning between white-beaked and common dolphins around the United Kingdom and Republic of Ireland. *Journal of the Marine Biological Association of the United Kingdom*, **88**, 1193-1198.

MacLeod, C.D., Weir, C.R., Pierpoint, C., Harland, E.J., 2007. The habitat preferences of marine mammals west of Scotland (UK). *Journal of the Marine Biological Association of the United Kingdom*, **87**, 157-164.

Macleod, K., Fairbairns, R., Gill, A., Fairbairns, B., Gordon, J., Blair-Myers, C., Parsons, E.C.M., 2004. Seasonal distribution of minke whales *Balaenoptera acutorostrata* in relation to physiography and prey off the Isle of Mull, Scotland. *Marine Progress Series*, **277**, 263-274.

McClellan, C.M., Brereton, T., Dell'Amico, F., Johns, D.G., Cucknell, A.-C., Patrick, S.C., Penrose, R., Ridoux, V., Solandt, J.L., Stephan, E., Votier, S.C., Williams, R., Godley, B.J., 2014. Understanding the distribution of marine megafauna in the English Channel region: Identifying key habitats for conservation within the busiest seaway on Earth. *PLoS ONE*, **9**, e89720.

Miller, P.I., Scales, K.L., Ingram, S.N., Southall, E.J., Sims, D.W., 2015. Basking sharks and oceanographic fronts: quantifying associations in the north-east Atlantic. *Functional Ecology*, **29**, 1099-1109.

Paxton, C.G.M., Scott-Hayward, L.A.S, Rexstad, E., 2014. Statistical approaches to aid the identification of Marine Protected Areas for minke whale, Risso's dolphin, white-beaked dolphin and basking shark. *Scottish Natural Heritage Commissioned Report No. 594.*

Perrow, M.R., Harwood, A.J.P., Skeate, E.R., Praca, E., Eglington, S.M., 2015. Use of multiple data sources and analytical approaches to derive a marine protected area for a breeding seabird. *Biological Conservation*, **191**, 729-738.

Pirotta, E., Thompson, P.M., Miller, P.I., Brookes, K.L., Cheney, B., Barton, T.R., Graham, I.M., Lusseau, D., 2014. Scale-dependent foraging ecology of a marine top predator modelled using passive acoustic data. *Functional Ecology*, **28**, 206-217.

Ramesh, K., Berrow, S., Meade, R., O'Brien, J., 2021. Habitat modelling on the potential impacts of shipping noise on fine whales (*Balaenoptera physalus*) in offshore Irish waters off the Porcupine Ridge. *Journal of Marine Science and Engineering*, **9**, 1207.

Robbins, J.R., Babey, L., Embling, C.B., 2020. Citizen science in the marine environment: estimating common dolphin densities in the north-east Atlantic. *PeerJ*, **8**, e8335.

Robinson, K.P., Tetley, M.J., Mitchelson-Jacob, E.G., 2009. The distribution and habitat preference of coastally occurring minke whales (*Balaenoptera acutorostrata*) in the outer southern Moray Firth, northeast Scotland. *Journal of Coastal Conservation*, **13**, 39-48.

Rogan, E., Cañadas, A., Macleod, K, Begoña Santos, M., Mikkelsen, B., Uriarte, A., Van Canneyt, O., Vásquez, 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.

Skov, H., Humphreys, E., Garthe S., Geitner, K., Grémillet, D., Hamer, K.C., Hennicke, J., Parner, H., Wanless, S., 2008. Application of habitat suitability modelling to tracking data of marine animals as a means of analyzing their feeding habitats. *Ecological Modelling*, **212**, 504-512.

Thompson, P.M., Brookes, K.L., Cordes, L.S., 2015. Integrating passive acoustic and visual data to model spatial patterns of occurrence in coastal dolphins. *ICES Journal of Marine Science*, **72**, 651-660.

Todd, N.R.E., Cronin, M., Luck, C., Bennison, A., Jessopp, A., Kavanagh, A.S., 2020. Using passive acoustic monitoring to investigate the occurrence of cetaceans in a protected marine area in northwest Ireland. *Estuarine, Coastal and Shelf Science*, **232**, 106509.

Waggitt, J.J., Cazenave, P.W., Torres, R., Williamson, B.J., Scott, B.E., 2016. Quantifying pursuit-diving seabirds' associations with fine-scale physical features in tidal stream environments. *Journal of Applied Ecology*, **53**, 1653-1666.

Warwick-Evans, V.C., Atkinson, P.W., Robinson, L.A., Green, J.A., 2016. Predictive modelling to identify near-shore, fine-scale seabird distributions during the breeding season. *PLoS ONE*, **11**, e0150592.

Williamson, L.D., Brookes, K.L., Scott, B.E., Graham, I.M., Bradbury, G., Hammond, P.S., Thompson, P.M., 2016. Echolocation detections and digital video surveys provide reliable estimates of the relative density of harbour porpoises. *Methods in Ecology and Evolution*, **7**, 762-769.

Wright, P.J., Begg, G.S., 1997. A spatial comparison of common guillemots and sandeels in Scottish waters. *ICES Journal of Marine Science*, **54**, 578-592.

Appendix 2. Species featured in the SDMs

Species	Scientific Name	Reference(s)
Atlantic Puffin	*Fratercula arctica*	Johnston *et al*. (2015); Waggitt *et al*. (2016)
Atlantic White-sided Dolphin	*Lagenorhynchus acutus*	Breen *et al*. (2016); MacLeod *et al*. (2007);
Auk spp.	Alcidae	McClellan *et al*. (2014)
Balaenoptera spp.	*Balaenoptera*	Baines *et al*. (2017)
Basking Shark	*Cetorhinus maximus*	Austin *et al*. (2019); McClellan *et al*. (2014); Miller *et al*. (2015); Paxton *et al*. (2015)
Beaked Whale spp.	Ziphiidae	Rogan *et al*. (2017)
Black Guillemot	*Cepphus grylle*	Waggitt *et al*. (2016)
Black-legged Kittiwake	*Rissa tridactyla*	Johnston *et al*. (2015)
Bottlenose Dolphin	*Tursiops truncates*	Arso Civil *et al*. (2019); Bailey & Thompson (2009); Breen *et al*. (2016); Pirotta *et al*. (2014)
Common Dolphin	*Delphinus delphis*	Breen *et al*. (2017); MacLeod *et al*. (2007); MacLeod *et al*. (2008b); Robbins *et al*. (2020)
Common Guillemot	*Uria aalge*	Johnston *et al*. (2015); Waggitt *et al*. (2016); Warwick-Evans *et al*. (2016); Wright & Begg (1997)
Dolphin spp.	Delphinidae	Cox *et al*. (2017); Hastie *et al*. (2005); McClellan *et al*. (2014); Thompson *et al*. (2015); Todd *et al*. (2020)
European Shag	*Phalacrocorax aristotelis*	Grémillet *et al*. (2020); Waggitt *et al*. (2016); Warwick-Evans *et al*. (2016)
Fin Whale	*Balaenoptera physalus*	Breen *et al*. (2016); Ramesh *et al*. (2021)
Great Black-backed Gull	*Larus marinus*	Johnston *et al*. (2015); Warwick-Evans *et al*. (2016)

Species	Scientific Name	Reference(s)
Grey Seal	*Halichoerus grypus*	Aarts *et al*. (2008); Bailey & Thompson (2009); Huon *et al*. (2015); Lieber *et al*. (2018); MacLeod *et al*. (2007);
Harbour Porpoise	*Phocoena phocoena*	Bailey & Thompson (2009); Breen *et al*. (2017); Brookes *et al*. (2013); Cox *et al*. (2017); de Boer *et al*. (2014); Embling *et al*. (2010); Gilles *et al*. (2016); Hammond *et al*. (2013); Isojunno *et al*. (2012); MacLeod *et al*. (2007); MacLeod *et al*. (2008a); McClellan *et al*. (2014); Todd *et al*. (2020); Williamson *et al*. (2016)
Harbour Seal	*Phoca vitulina*	Bailey & Thompson (2009); Bailey *et al*. (2014); Hastie *et al*. (2018); Jones *et al*. (2017); Lieber *et al*. (2018)
Herring Gull	*Larus argentatus*	Warwick-Evans *et al*. (2016)
Leatherback Turtle	*Dermoxhelys coriacea*	McClellan *et al*. (2014)
Lesser Black-backed Gull	*Larus fuscus*	Johnston *et al*. (2015); Warwick-Evans *et al.* (2016)
Little Auk	*Alle alle*	Johnston *et al*. (2015)
Little Tern	*Sternula albifrons*	Perrow *et al*. (2015)
Long-finned Pilot Whale	*Globicephala melas*	Breen *et al*. (2016); MacLeod *et al*. (2007); Rogan *et al*. (2017)
Manx Shearwater	*Puffinus puffinus*	Kane *et al*. (2020)
Minke Whale	*Balaenoptera acutorostrata*	Anderwald *et al*. (2012); Breen *et al*. (2016); Hammond *et al*. (2013); Macleod *et al*. (2004); MacLeod *et al*. (2007); Paxton *et al*. (2014); Robinson *et al*. (2009)
Northern Fulmar	*Fulmarus glacialis*	Johnston *et al*. (2015)
Northern Gannet	*Morus bassanus*	Johnston *et al*. (2015); McClellan *et al*. (2014); Skov *et al*. (2008)
Razorbill	*Alca torda*	Johnston *et al*. (2015); Warwick-Evans *et al.* (2016)

Species	Scientific Name	Reference(s)
Risso's Dolphin	*Grampus griseus*	Breen *et al*. (2016); de Boer *et al.* (2014); Paxton *et al*. (2014)
Sperm Whale	*Physeter macrocephalus*	Breen *et al.* (2016); Rogan *et al*. (2017)
White-beaked Dolphin	*Lagenorhynchus albirostris*	Breen *et al*. (2016); MacLeod *et al*. (2007); MacLeod *et al*. (2008b); Paxton *et al.* (2014)

Appendix 3. Composition of environmental predictor variables featured in this report that consist of more than one term identified in the literature.

Variable	Description
Anthropogenic Noise/Activity	Boat speed, number of shipping tonal detections, playback status, presence/absence of construction activity, remote ship noise, seismic ship noise, shipping noise level, survey vessel noise, water noise level
Change in Tide Height	Mean difference between high and low water at nearest harbour, rate of change in tide
Current Speed	Current speed, mean relative variance in velocity, peak flow, spatial variation of current speed, vertical current speed
Distance to Bathymetric Contour	Distance to 200m isobath, distance to 2000m isobath
Rugosity	Seafloor rugosity, seabed roughness, contour index (*defined as a measure of variability in the seabed*), standard deviation of depth
Salinity	Sea surface salinity, sea bottom salinity
Sea Surface Height	Sea surface height, water elevation
Spring-Neap Tide Cycle	Days before/after neap tide, position in spring-neap cycle
Tidal Stratification	Tidal stratification, mean stratification, tidal mixing
Time of Day	Daytime, night-time, time after sunrise, time of day, time to sunset

Bibliography

BENTLAGE, B., PETERSON, A.T., BARVE, N., and CARTWRIGHT, P., 2013. Plumbing the depths: extending ecological niche modelling and species distribution modelling in three dimensions. *Global Ecology and Biogeography*, **22**, 952-961.

BOLTON, M., 2020. GPS tracking reveals highly consistent use of restricted foraging areas by European Storm-petrels *Hydrobates pelagicus* breeding at the largest UK colony: implications for conservation management. *Bird Conservation International*, 1-18.

BREEN, P., BROWN, S., REID, D., and ROGAN, E., 2017. Where is the risk? Integrating a spatial distribution model and a risk assessment to identify areas of cetacean interaction with fisheries in the northeast Atlantic. *Ocean & Coastal Management*, **136**, 148-155.

BURKE, C.M., and MONTEVECCHI, W.A., 2009. The foraging decisions of a central place foraging seabird in response to fluctuations in local prey conditions. *Journal of Zoology*, **278**, 354-361.

COX, S.L., EMBLING, C.B., HOSEGOOD, P.J., VOTIER, S.C., and INGRAM, S.N., 2018. Oceanographic drivers of marine mammal and seabird habitat-use across shelf-seas: A guide to key features and recommendations for future research and conservation management. *Estuarine, Coastal and Shelf Science*, **212**, 294-310.

DIAS, M.P., MARTIN, R., PEARMAIN, E.J., BURFIELD, I.J., SMALL, C., PHILLIPS, R.A., YATES, O., LASCELLES, B., BORBOROGLU, P.G., and CROXALL, J.P., 2019. Threats to seabirds: A global assessment. *Biological Conservation*, **237**, 525-537.

ELITH, J., GRAHAM, C.H., ANDERSON, R.P., DUDÍK, M., FERRIER, S., GUISAN, A., HIKMANS, R.J., HUETTMANN, F., LEATHWICK, J.R., LEHMANN A., LI, J., LOHMANN, L.G., LOISELLE, B.A., MANION, G., MORITZ, C., NAKAMURA, M., NAKAZAWA, Y., OVERTON, J.M., PETERSON, A.T., PHILLIPS, S.J., RICHARDSON, K., SCACHETTI-PEREIRA, R., SCHAPRE, R.E., SOBERÓN, J., WILLIAMS, S., WISZ M.S., and ZIMMERMANN, N.E., 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151.

ELITH, J., and LEATHWICK, J.R., 2009. Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 677-697.

GUISAN, A., and ZIMMERMANN, N.E., 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147-186.

HASTIE, T., and TIBSHIRANI, R., 1986. Generalized Additive Models. *Statistical Science*, **1**, 297-318.

HOEGH-GULDBERG, O., and BRUNO, J.F., 2010. The impact of climate change on the world's marine ecosystems. *Science*, **328**, 1523-1528.

KEARNEY, M., and PORTER, W., 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, **12**, 334-350.

MCCLELLAN, C.M., BRERETON, T., DELL'AMICO, F., JOHNS, D.G., CUCKNELL, A.-C., PATRICK, S.C., PENROSE, R., RIDOUX, V., SOLANDT, J.L., STEPHAN, E., VOTIER, S.C., WILLIAMS, R., and GODLEY, B.J., 2014. Understanding the distribution of marine megafauna in the English Channel region: Identifying key habitats for conservation within the busiest seaway on Earth. *PLoS ONE*, **9**, e89720.

MCCONNELL, B.J., FEDAK, M.A., LOWELL, P., and HAMMOND, P.S., 1999. Movements and foraging areas of grey seals in the North Sea. *Journal of Applied Ecology*, **36**, 573-590.

MELO-MERINO, S.M., REYES-BONILLA, H., and LIRA-NORIEGA, A., 2020. Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. *Ecological Modelling*, **415**, 108837.

NAUD, M.-J., LONG, B., BRÊTHES, J.-C., and SEARS, R., 2003. Influences of underwater bottom topography and geomorphology on minke whale (*Balaenoptera acutorstrata*) distribution in the Mingan Islands (Canada). *Journal of the Marine Biological Association of the United Kingston*, **83**, 889-896.

NELDER, J.A., and WEDDERBURN, R.W.M., 1972. Generalized Linear Models. J*ournal of the Royal Statistical Society. Series A (General)*, **135**, 370-384.

NOWACEK, D.P., CHRISTIANSE, F., BEJDER, L., GOLDBOGEN, J.A., and FRIEDLAENDER, A.S., 2016. Studying cetacean behaviour: new technological approaches and conservation applications. *Animal Behaviour*, **120**, 235-244.

PETERSON, A.T., PAPES, M., and EATON, M., 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography*, **30** 550-560.

PHILLIPS, S.J., ANDERSON, R.P., and SCHAPIRE, R.E., 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231-259.

QUILLFEDLT, P., MICHALIK, A, VEIT-KÖHLER, G., STRANGE, I.J., and MASELLO, J.F., 2010. Inter-annual changes in diet and foraging trip lengths in a small pelagic seabird, the thin-billed prion *Pachyptila belcheri*. *Marine Biology*, **157**, 2043-2050.

REDFERN, J.V., FERGUSON, M.C., BECKER, E.A., HYRENBACH, K.D., GOOD, C., BARLOW, J., KASCHNER, K., BAUMGARTNER, M.F., FORNEY, K.A., BALANCE, L.T., FAUCHALD, P., HALPIN, P., HAMAZAKI, T., PERSHING, A.J., QIAN, S.S., READ, A., REILLY, S.B., TORRES, L., and WERNER, F., 2006. Techniques for cetacean-habitat modeling. *Marine Ecology Progress Series*, **310**, 271-295.

ROBINSON, L.M., ELITH, J., HOBDAY, A.J., PEARSON, R.G., KENDALL, B.E., POSSINGHAM, H.P., and RICHARDSON, A.J., 2011. Pushing the limits in marine species distribution modelling: lessons from the land present challenges and opportunities. *Global Ecology and Biogeography*, **20**, 789-802.

ROGAN, E., CAÑADAS, A., MACLEOD, K., SANTOS, M.B., MIKKELSEN, B., URIARTE, A., VAN CANNEYT, O., VÁSQUEZ, J.A., and 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.

SCALES, K.L., HAZEN, E.L., JACOX, M.G., EDWARDS, C.A., BOUSTANY, A.M., OLIVER, M.J., and BOGRAD, S.J., 2017. Scale of inference: on the sensitivity of habitat models for wide-ranging marine predators to the resolution of environmental data. *Ecography*, **40**, 210-220.

SMOLIK. M.G., DULLINGER, S., ESSL, F., KLEINBAUER, I., LEITNER, M., PETERSEIL, J., STADLER, L.-M., and VOGL, G., 2010. Integrating species distribution models and interacting particle systems to predict the spread of an invasive alien plant. *Journal of Biogeogrpahy*, **37**, 411-422.

STOCKWELL, D., and PETERS, D., 1999. The GARP modelling system: problems and solution to automated spatial prediction. *International Journal of Geographical Information Science*, **13**, 143-158.

THUSHARI, G.G.N., and SENEVIRATHNA, J.D.M., 2020. Plastic pollution in the marine environment. *Heliyon*, **6**, e04709.

TREMBLAY, Y., BERTRAND, S., HENRY, R.W., KAPPES, M.A., COSTA, D.P., and SHAFFER, S.A., 2009. Analytical approaches to investigating seabird-environment interactions: a review. *Marine Ecology Progress Series*, **391**, 153-163.

VAN DER KOOIJ, J., SCOTT, B.E., and MACKINSON, S., 2008. The effects of environmental factors on daytime sandeel distribution and abundance on the Dogger Bank. *Journal of Sea Research*, **60**, 201-209.

ŽYDELIS, R., WALLACE, B.P., GILMAN, E.L., and WERNER, T.B., 2009. Conservation of marine megafauna through minimization of fisheries bycatch. *Conservation Biology*, **23**, 608-616.

List of abbreviations

EENM – Ensemble Ecological Niche Model

ENFA – Ecological Niche Factor Analysis

GAM – Generalised Additive Model

GAMM – Generalised Additive Mixed Model

GARP – Genetic Algorithm for Rule Set Production

GLM – Generalised Linear Model

GLMM – Generalised Linear Mixed Model

PCA – Principal Component Analysis

SDM – Species Distribution Model

Natural England is here to secure a healthy natural environment for people to enjoy, where wildlife is protected and England's traditional landscapes are safeguarded for future generations.

Natural England publications are available as accessible pdfs from www.gov.uk/natural-england.

Should an alternative format of this publication be required, please contact our enquiries line for more information: 0300 060 3900 or email enquiries@naturalengland.org.uk.

Catalogue code: NECR447