RESEARCH ARTICLE

Distribution maps of cetacean and seabird populations in the North-East Atlantic

James J. Waggitt[1] | **Peter G. H. Evans**[1,2] | **Joana Andrade**[3] | **Alex N. Banks**[4] | **Oliver Boisseau**[5] | **Mark Bolton**[6] | **Gareth Bradbury**[7] | **Tom Brereton**[8] | **Cornelis Jan Camphuysen**[9] | **Jan Durinck**[10] | **Tom Felce**[11] | **Ruben Christiaan Fijn**[12] | **Isabel Garcia-Baron**[13] | **Stefan Garthe**[14] | **Steve C. V. Geelhoed**[15] | **Anita Gilles**[16] | **Martin Goodall**[17] | **Jan Haelters**[18] | **Sally Hamilton**[19] | **Lauren Hartny-Mills**[20] | **Nicola Hodgins**[21] | **Kathy James**[2] | **Mark Jessopp**[22,23] | **Ailbhe S. Kavanagh**[22] | **Mardik Leopold**[15] | **Katrin Lohrengel**[2] | **Maite Louzao**[13] | **Nele Markones**[14] | **Jose Martínez-Cedeira**[24] | **Oliver Ó Cadhla**[25] | **Sarah L. Perry**[26] | **Graham J. Pierce**[27] | **Vincent Ridoux**[28] | **Kevin P. Robinson**[29] | **M. Begoña Santos**[30] | **Camilo Saavedra**[30] | **Henrik Skov**[31] | **Eric W. M. Stienen**[32] | **Signe Sveegaard**[33] | **Paul Thompson**[34] | **Nicolas Vanermen**[32] | **Dave Wall**[35] | **Andy Webb**[36] | **Jared Wilson**[37] | **Sarah Wanless**[38] | **Jan Geert Hiddink**[1]

[1]School of Ocean Sciences, Bangor University, Menai Bridge, UK; [2]Sea Watch Foundation, Amlwch, UK; [3]Sociedade Portuguesa para o Estudo das Aves, Lisboa, Portugal; [4]Natural England, Exeter, UK; [5]Marine Conservation Research, Kelvedon, UK; [6]Royal Society for the Protection of Birds Centre for Conservation Science, Sandy, UK; [7]Wildfowl and Wetlands Trust Consulting, Slimbridge, UK; [8]Marinelife, Bridport, UK; [9]Royal Netherlands Institute for Sea Research, Texel & Utrecht University, Texel, Netherlands; [10]Marine Observers, Snedsted, Denmark; [11]Manx Whale and Dolphin Watch, Peel, Isle of Man; [12]Bureau Waardenburg, Culemborg, Netherlands; [13]AZTI Fundazioa, Pasaia, Spain; [14]Research and Technology Centre (FTZ), University of Kiel, Büsum, Germany; [15]Wageningen Marine Research, Den Helder, The Netherlands; [16]Institute for Terrestrial and Aquatic Wildlife Research, University of Veterinary Medicine Hannover Foundation, Büsum, Germany; [17]Cornwall Wildlife Trust, Truro, UK; [18]Royal Belgian Institute of Natural Sciences, Ostend, Belgium; [19]ORCA, Portsmouth, UK; [20]Hebridean Whale and Dolphin Trust, Isle of Mull, UK; [21]Whale and Dolphin Conservation, Chippenham, UK; [22]MaREI Centre, Environmental Research Institute, University College Cork, Cork, Ireland; [23]School of Biological, Earth & Environmental Sciences, University College Cork, Cork, Ireland; [24]Coordinadora para o Estudo dos Mamíferos Mariños, Pontevedra, Spain; [25]Science and Biodiversity Section, National Parks & Wildlife Service, Galway, Ireland; [26]Cardigan Bay Marine Wildlife Centre, The Wildlife Trust of South and West Wales, New Quay, UK; [27]Instituto de Investigaciones Marinas (CSIC), Vigo, Spain; [28]Observatoire PELAGIS, UMS 3462, La Rochelle Université/CNRS, La Rochelle, France; [29]Cetacean Research & Rescue Unit, Banff, UK; [30]Instituto Español de Oceanografía, Centro Oceanográfico de Vigo, Vigo, Spain; [31]DHI, Hørsholm, Denmark; [32]Research Institute for Nature and Forest, Brussels, Belgium; [33]Department of Bioscience, Aarhus University, Roskilde, Denmark; [34]Lighthouse Field Station, University of Aberdeen, Cromarty, UK; [35]Irish Whale and Dolphin Group, Kilrush, Ireland; [36]HiDef Aerial Surveying Ltd., Cumbria, UK; [37]Marine Scotland Science, Aberdeen, UK and [38]Centre for Ecology & Hydrology, Penicuik, UK

Correspondence
James J. Waggitt
Email: j.waggitt@bangor.ac.uk

Funding information
Marine Ecosystems Research Programme, Grant/Award Number: NE/L003201/1

Handling Editor: Andre Punt

Abstract

1. Distribution maps of cetaceans and seabirds at basin and monthly scales are needed for conservation and marine management. These are usually created from standardized and systematic aerial and vessel surveys, with recorded animal densities interpolated across study areas. However, distribution maps at basin and monthly scales have previously not been possible because individual surveys have restricted spatial and temporal coverage.

2. This study develops an alternative approach consisting of: (a) collating diverse survey data to maximize spatial and temporal coverage, (b) using detection functions to estimate variation in the surface area covered (km²) among these surveys,

253

standardizing measurements of effort and animal densities, and (c) developing species distribution models (SDM) that overcome issues with heterogeneous and uneven coverage.

3. 2.68 million km of survey data in the North-East Atlantic between 1980 and 2018 were collated and standardized. SDM using Generalized Linear Models and General Estimating Equations in a hurdle approach were developed. Distribution maps were then created for 12 cetacean and 12 seabird species at 10 km and monthly resolution. Qualitative and quantitative assessment indicated good model performance.

4. *Synthesis and applications*. This study provides the largest ever collation and standardization of diverse survey data for cetaceans and seabirds, and the most comprehensive distribution maps of these taxa in the North-East Atlantic. These distribution maps have numerous applications including the identification of important areas needing protection, and the quantification of overlap between vulnerable species and anthropogenic activities. This study demonstrates how the analysis of existing and diverse survey data can meet conservation and marine management needs.

KEYWORDS

Bay of Biscay, Celtic Sea, detection function models, English Channel, Hebrides, Irish Sea, North Sea, species distribution models

1 | INTRODUCTION

Environmental change (Hoegh-Guldberg & Bruno, 2010) and anthropogenic activities (Halpern et al., 2015, 2008) can have profound impacts on marine ecosystems. In many cases, assessing these impacts requires an understanding of species distributions. For instance, knowing species distributions identifies the proportion of populations interacting with anthropogenic activities, explaining declines (Boivin et al., 2016) and/or helping develop appropriate mitigation and management solutions (Wood, 2003). Information on species distributions at monthly and basin scales is needed in marine ecosystems, where large numbers of species routinely move hundreds or thousands of kilometres in migratory or dispersive movements (Hays & Scott, 2013).

As apex predators, cetaceans and seabirds have important ecological roles including the top-down regulation of lower trophic levels (Hunt & McKinnell, 2006) and the transport of nutrients (Doughty et al., 2016). They are also charismatic species of socio-economic importance, due to their cultural appeal and focus for eco-tourism (Higham & Lück, 2007). However, these taxa face numerous anthropogenic threats including bycatch, habitat-loss, energy extraction, noise disturbance, prey reductions, pollution and vessel traffic (Avila, Kaschner, & Dormann, 2018; Croxall et al., 2012). Since their conservation is of importance for regulatory bodies, the need for distribution maps at monthly and basin scales has been recognized by the European Union (Habitats Directive:

92/43/EEC, Birds Directive: 2009/147/EC, Marine Strategy Framework Directive: 2008/56/EC).

Distribution maps of cetaceans and seabirds are usually produced from transects using humans/cameras on moving platforms to record animals (Buckland et al., 2012; Camphuysen, Fox, Leopold, & Petersen, 2004; Evans & Hammond, 2004). Animal densities (individuals per km^2) are then estimated along transects (Buckland et al., 2001), before being interpolated across study areas (Hammond et al., 2013). In most cases, transects are performed using similar platforms and observation methods, providing comparable measurements of surface area covered and animal densities. Systematic transect designs are also used, providing homogeneous and even coverage. However, due to financial and logistical constraints, surveys covering whole basins occur at decadal intervals (Hammond et al., 2002, 2013) while those covering seasonal cycles focus on relatively small areas (Gilles et al., 2016). Therefore, distribution maps at monthly and basin scales are lacking, and their provision demands an alternative approach.

This study develops an alternative approach to provide distribution maps for 12 cetacean and 12 seabird species (Table 1) at 10 km and monthly resolution in the North-East Atlantic. This approach consists of three stages. First, effort in time and space is maximized by collating survey data from as many different sources and suppliers as possible (Mannocci et al., 2018; Paxton, Scott-Hayward, Mackenzie, Rexstad, & Thomas, 2016; Roberts et al., 2016). Second, differences among surveys linked with platform-type (aircraft vs. vessel, low vs. high), transect-design (line-transect vs. strip-transect),

TABLE 1 A summary of the cetacean and seabird species analysed in this study including their identification code, detection group, and months of nest-occupancy (for seabirds)

Taxa	Common name	Scientific name	Code	Group	Nest
Cetacean	Atlantic White-Sided Dolphin	*Lagenorhynchus acutus*	AWSD	A	—
	Bottlenose Dolphin	*Tursiops truncatus*	BND	A	—
	Fin Whale	*Balaenoptera physalus*	FW	C	—
	Harbour Porpoise	*Phocoena phocoena*	HP	B	—
	Killer Whale	*Orcinus orca*	KW	D	—
	Long-Finned Pilot Whale	*Globicephala melas*	LFPW	D	—
	Minke Whale	*Balaenoptera acutorostrata*	MW	E	—
	Risso's Dolphin	*Grampus griseus*	RD	D	—
	Short-Beaked Common Dolphin	*Delphinus delphis*	SBCD	A	—
	Sperm Whale	*Physeter macrocephalus*	SPW	F	—
	Striped Dolphin	*Stenella coeruleoalba*	SD	A	—
	White-Beaked Dolphin	*Lagenorhynchus albirostris*	WBD	A	—
Seabird	Atlantic Puffin	*Fratercula arctica*	PUF	J	Apr–Aug
	Black-Legged Kittiwake	*Rissa tridactyla*	KIT	M	Apr–Aug
	Common Guillemot	*Uria aalge*	GIL	J	Apr–Jul
	European Shag	*Phalacrocorax aristotelis*	SHG	O	Mar–Aug
	European Storm Petrel	*Hydrobates pelagicus*	ESP	G	May–Sep
	Great Skua	*Stercorarius skua*	GRK	K	Apr–Jul
	Herring Gull	*Larus argentatus*	HEG	L	Apr–Jul
	Lesser Black-Backed Gull	*Larus fuscus*	LBB	L	Apr–Jul
	Manx Shearwater	*Puffinus puffinus*	MSH	N	Apr–Aug
	Northern Fulmar	*Fulmarus glacialis*	FUL	H	Apr–Aug
	Northern Gannet	*Morus bassanus*	GAN	I	Apr–Sep
	Razorbill	*Alca torda*	RAZ	J	Apr–Jul

observation method (human vs. camera) and weather (sea state) are accounted for by calculating variations in the surface area covered (Buckland et al., 2001). Finally, species distribution models (SDM; Elith & Leathwick, 2009) are used to overcome problems with heterogeneous and uneven coverage in collations of survey data (Paxton et al., 2016).

2 | MATERIALS AND METHODS

2.1 | Collation

Aerial and vessel survey data were collated from the North-East Atlantic between 1980 and 2018. The North-East Atlantic was considered here to represent areas spanning between Norway and Iberia on a north–south axis, and Rockall to the Skagerrak on an east–west axis. Only survey data collected using dedicated human observers (i.e. not performing other duties) or cameras to record animals were used. Survey data also needed to include information for the calculation of variations in the surface area covered among surveys; namely platform-type, platform-height, transect-design and recording method. Survey data were screened

for typographical and positional errors. Platforms and sightings recorded as being on land (i.e. incorrect coordinates) were removed. Platforms recorded as travelling at unrealistic speeds were also removed. To do so, mean (μ) speeds were calculated for each platform. For each vessel, speeds greater than $\mu + \mu/2$ were then removed. For each aircraft, those less than $\mu - \mu/4$ or greater than $\mu + \mu/4$ were removed. These differences were because vessels but not aircraft can move at low speeds.

2.2 | Standardization

The surface area covered is described using a perpendicular distance from the transect line, and is commonly referred to as the effective strip width (*esw*). The *esw* differs between line and strip-transects. In the latter, observations focus up to a pre-defined distance. It is assumed that all animals in this area are detected. This distance represents the *esw*. In the former, observations focus on all distances. It is assumed that the detection of animals decreases with increasing distance. Therefore, distances between animals and transect lines are recorded, and these distances are used to estimate the *esw*. An intermediate method (European Seabirds At Sea: ESAS) also exists for

cetaceans and seabirds on the water whereby observations focus up to a pre-defined distance, but distances to animals are recorded into a series of distance bands (Camphuysen et al., 2004). Strip-transects have either human or camera observations, whereas line-transects and ESAS have only human observations. Surveys commonly use a combination of transect designs with cetaceans, seabirds on the water, and seabirds in flight recorded differently.

2.2.1 | Line and ESAS transects

Variations in *esw* among surveys using line-transects and ESAS were estimated using detection function models (Buckland et al., 2001). Different models were developed for each combination of species, survey method (line-transect vs. ESAS), behaviour (on the water surface or in flight) and platform (vessel vs. aircraft). This approach accounted for differences in the factors influencing detectability of animals among these categories. As with previous studies (Paxton et al., 2016), species were grouped together based upon their morphological and behavioural traits (Table 1). As morphology and behaviour affects detectability, group members were assumed to have identical detectability. This grouping increased sample sizes for detection function models, and provided a broader range of scenarios for estimation of variations in *esw* among surveys. For instance, if a particular survey method or platform dominated the core-range of a particular species, then reliable estimations of *esw* for other survey methods or platforms would not be possible. The perpendicular distance between the transect line and animals (m) was the response variable. Distances to animals were recorded for most relevant sightings (cetaceans = 78%, seabirds on the water = 70%, seabirds in flight = 99%). The central distance of bands were used for ESAS while absolute distances were used for line-transects. Platform height (observer height above sea surface, m) and sea state (Beaufort scale) were explanatory variables (Table 2), and modelled as continuous variables. As precise information on platform height was not always available, heights were assigned to discrete categories, with the central height used (Table 2). Values of platform height and sea state were log-transformed, as the influence of increasing values would be greatest among smaller vessels and lower sea states. Additional factors influencing the detection of animals were not included because they were recorded in an inconsistent manner (weather), highly subjective (observer experience) or collinear with platform height (vessel speed).

All combinations of explanatory variables were tested, and both half-normal and hazard-rate responses were trialled. The detection function was truncated at the pre-defined distance for ESAS and at 1 km for line-transects. The latter was because sightings beyond 1 km were rare (cetaceans = 3%, seabirds = <1%). Positive relationships between *esw* and sea state seem unlikely, and presumably arise when the core-range of a particular species coincides with surveys experiencing rougher weather (i.e. those beyond the continental shelf edge). Therefore, combinations producing such relationships were ignored. Only survey data collected in sea state's of Beaufort

TABLE 2 The explanatory variables used in detection functions estimating variations in effective strip width (*esw*) and probability of detection on the track-line (*g*(0))

Variable	Type	Measure	Description
Platform	Continuous	2.5 m	Vessels with observers at 0–2.5 m above sea level
		5 m	Vessels with observers at 2.5–10 m above sea level
		10 m	Vessels with observers at 5–10 m above sea level
		20 m	Vessels with observers at 10–20 m above sea level
		30 m	Vessels with observers at 20–30 m above sea level
		75 m	Aircraft with observers at 50–100 m above sea level
		150 m	Aircraft with observers at 100–200 m above sea level
Sea state	Continuous	0.5 to 3	Beaufort scale

scale 3 or less were considered, to ensure that only those collected in good conditions contributed to analyses. The model producing the lowest Akaike's Information Criteria was used to estimate variations in *esw* among species and surveys. Detection function models were fitted using the package 'MRDS' (Thomas et al., 2010) in R (v.3.2.5, R Development Core Team, 2016).

2.2.2 | Strip-transects

Variations in *esw* among surveys using strip-transects (both human and camera observations) were determined using information provided from data suppliers.

2.2.3 | Adjustments to esw

The calculation of *esw* assumes that the probability of detecting animals on the transect line, commonly known as *g*(0), equals 1. However, in surveys using observers, *g*(0) varies greatly due to biases (Buckland et al., 2001). Perception bias describes where observers miss animals because their visibility is compromised, perhaps due to high sea state. Availability bias describes when observers miss animals because they are undetectable, usually because cetaceans and diving seabirds (Alcidae, European shag, Manx shearwater) are below the water surface. Finally, response bias describes where animals react to the presence of the platform. For example, dolphins often approach vessels, harbour porpoises move away from vessels, and scavenging seabirds (Laridae, northern gannet, northern fulmar) follow vessels. These biases could differ among platforms and sea state. However, ignoring them can produce misleading estimations

of densities by under or overestimating the *esw* for a particular scenario or species (Hammond, 2010).

For vessel surveys, it was assumed that all biases were relevant. These biases are collectively accounted for using a double-platform survey with primary and secondary observers. The secondary observers focus on the track line further ahead of the vessel. They aim to detect animals before responsive movement. Estimation of *g*(0) is possible by comparing the sightings of the primary and secondary observers, (Burt, Borchers, Jenkins, & Marques, 2014). Unfortunately double-platform surveys were absent for seabirds, meaning that variations in *g*(0) among vessel surveys could not be estimated. However, 77,570 km of double-platform surveys were available for cetaceans, enabling these variations to be estimated using a full-independence mark-recapture model (Burt et al., 2014). As with previous studies (Paxton et al., 2016), estimations of variation in *g*(0) across platform height and sea state allow predictions on occasions where double-platform surveys were not used, increasing the compatibility of these surveys. The presence/absence of a resighting by the primary observer was the response variable. Log-transformed values of platform height and sea state were explanatory variables. Selection and predictions from optimal models followed procedures for *esw*. Models were fitted using the package 'MRDS' in R.

For aerial surveys, it was assumed that only availability bias was relevant. Availability bias was considered trivial for diving seabirds, as animals are usually visible (Thaxter et al., 2010; Wanless, Corfield, Harris, Buckland, & Morris, 1993). However, availability biases were considered non-trivial for cetaceans, as animals are mainly underwater. *g*(0) for cetaceans was represented by the proportion of time that animals spend at the sea surface. These approaches are admittedly simplistic; availability bias could depend on observation technique (fixed or scanning) in combination with aircraft speed, whilst perception bias is considered likely (Borchers, Zucchini, Heide-Jørgensen, Cañadas, & Langrock, 2013). However, robust estimation of *g*(0) across scenarios (survey method, platform height and sea state) were neither available nor achievable from relevant sightings. Information on the proportion of time that animals spend at the sea surface were sourced from previous studies (Alves et al., 2013; Hansen et al., 2018; Rasmussen, Akamatsu, Teilmann, Vikingsson, & Miller, 2013; Watwood, Miller, Johnson, Madsen, & Tyack, 2006).

Final calculations

The surface area covered (km^2) per transect was calculated using Equation 1: *L* is the transect length (km) and *s* is the number of platform sides covered by observers (1 or 2).

$$\text{Area Searched} = esw\, g(0)\, s\, L. \tag{1}$$

2.3 | Species distribution models

Spatial and temporal variations in species presence (0 = absent, 1 = present), animal density (individuals per km^2), the surface area covered (km^2), and environmental conditions were quantified in a 10 km resolution orthogonal grid. These measurements were provided for each combination of platform, day, and cell. For seabirds, two measurements of the surface area covered and animal densities were provided—one for those on the sea surface, and one for those in flight. The final measurement of animal densities represented the sum of these components. Transects were split at cell boundaries when they spanned several cells. Processing was performed using the 'RASTER' package (Hijmans, 2013) in R.

2.3.1 | Sightings

There are profound ecological differences between coastal and offshore bottlenose dolphin *Tursiops truncatus* (Hoelzel, Potter, & Best, 1998; Louis et al., 2014). This study focussed on the offshore ecotype to avoid confounding influences hindering the development of SDM for either ecotype, and because the distribution of the coastal ecotype is relatively well known (Reid, Evans, & Northridge, 2003). Bottlenose dolphins encountered more than 30 km from the coastline were considered to represent the offshore ecotype (Breen, Brown, Reid, & Rogan, 2016). Discrimination between Alcidae (common guillemot *Uria aalge*, razorbill *Alca torda*) species is often difficult, particularly in aerial and digital surveys where observations are made at considerable altitude (Buckland et al., 2012). Discrimination between species was not possible in 37% of sightings, leading to underestimates of densities. Therefore, these sightings were assigned to species, based upon the relative proportion of each species in vessels surveys performed within 100 km in the same month. This distance was based upon the scale of their movements whilst resident in a region (Thaxter et al., 2012). No other modifications were made to the sightings data. While there is often uncertainty in the estimation of group-sizes for species forming large pods or flocks, lower and upper estimates were not provided by the vast majority of data suppliers. Therefore, it was not possible to account for any systematic variation in the misestimation of group sizes across survey method, platform height or sea state.

2.3.2 | Environmental conditions

Because this study aimed to produce distribution maps at basin and monthly scales, environmental conditions needed to discriminate among consistently different habitats (e.g. shallow vs. deep, warm vs. cool) and seasons (e.g. coolest vs. warmest months). Therefore, survey data were combined with average conditions for that month across years rather than concurrent conditions. Values of temperature (°C) were sourced from a FOAM AMM7 model available from the Marine Environmental Monitoring Systems (http://marine.copernicus.eu), providing values at 7 km and 1-month resolution at 30 depth intervals between 1985 and 2018. Values of seabed depth (m) were sourced from the EMODnet archive, and were provided at approximately 1 km resolution (http://www.emodnet-bathymetry.eu). Values of depth and temperature were then resampled at 10 km resolution using block-averaging and bilinear interpolation, respectively. In total, six environmental conditions were derived

from values of depth and temperature. Details on their calculation are summarized in Table 3. Spatial and temporal conditions rather than a single spatiotemporal condition were calculated from values of temperature. This choice was based on the concept that biogeographical ranges are determined by spatial variations in annual temperature, while seasonal movement around this range is a response to temporal variations in basin temperature.

Seabirds breed on land during the summer months. During this time they function as central place foragers, with distributions of species centred on large colonies (Gaston, 2004). To quantify the influence of colony location and size, a colony index was calculated for each species. To isolate the influence of colonies, these indices aimed to reproduce a scenario where animals dispersed evenly around a particular colony, and where the numbers of animals encountered decreased exponentially with increasing distance from this colony (Grecian et al., 2012). National censuses including locations and counts of breeding birds were obtained from nine countries (see Table S1). While the census was performed in different years, relatively large colonies (e.g. those in northern UK) should persist across the study period. Each cell containing breeding birds was considered as a colony. A colony-specific index ($COLs$) was first calculated for each cell in the study area. For each cell, the distance to the focal colony (km), the number of cells upto that distance to the focal colony (n), and the number of animals breeding in the focal colony (Pop) were calculated. The calculation of n excluded cells occurring on landmasses. In colonies where numbers of breeding birds were available for multiple years, Pop represented the mean number. In combination, these components were used in formula

2 to estimate how many animals would be expected in each cell given the scenario above ($COLs$).

$$COLs = \frac{Pop}{n} \tag{2}$$

This process was repeated for each colony in the study area, before a cumulative colony index (COL) was then calculated for each cell using formula 3.

$$COL = \sum COLs. \tag{3}$$

COL was then standardized between values of 0 and 1. This conversion means that COL merely describes the proximity of a cell to breeding aggregations, rather than animal densities on the assumption of even dispersal. This is particularly important for Laridae where many animals exploit terrestrial rather than marine environments (Kubetzki & Garthe, 2003). COL was weighted by whether survey data was during (1), within 1 month (0.5) or outside (0) the breeding season (Table 1). This final adjustment meant that high values of COL identified survey data that were collected near large breeding aggregations during the breeding season. All processing was performed using the 'RASTER' package (Hijmans, 2013) in R.

2.3.3 | Environmental associations

A hurdle approach was used to quantify associations between each species and environmental conditions. This approach comprises two

TABLE 3 The explanatory variables used in statistical models predicting spatial and temporal variations in animal densities

Variable	Type	Measure	Description	Source
Annual temperature	Spatial	°C	Mean temperature between 0 and 150 m depth[c]	FOAM AMM7 model
Annual temperature variance	Spatial	°C	Variance in temperature between 0 and 150 m depth[c]	FOAM AMM7 model
Breeding colony index	Spatial and temporal	Arbitrary	Proximity and size of nearest breeding colonies[a]	Various
Breeding cycle	Temporal	Arbitrary	Breeding season (1), 1-month side of either breeding season (0.5) or non-breeding season (0)[b]	Expert opinion
Depth	Spatial	m	Depth	EMODNet bathymetry
Fronts	Spatial	°C	Gradients in the prevalence of thermal stratification, calculated using the mean difference between the focal cell and its neighbouring cells. Thermal stratification is the absolute range in annual temperature (see above) between 0 and 150 m depth. Strong gradients indicate areas of intense fronts[c]	FOAM AMM7 model
Land	Spatial	Km	Distance to the nearest land mass	EMODNet bathymetry
Regional temperature	Temporal	°C	Mean temperature between 0 and 150 m depth during the month of the survey[c]	FOAM AMM7 model
Seabed roughness	Spatial	m	Gradients in depth, calculated using the mean difference between the focal cell and its neighbouring cells. Strong gradients indicate areas of uneven seabed including bank-systems, shelf-edges, slopes and trenches	EMODNet bathymetry

[a]See main text for calculations of breeding indices.

[b]See Table 1 for information on the breeding seasons of seabirds.

[c]Calculations used values between 1985 and 2018.

components: a presence–absence model relating to the probability of encountering animals, and a count model relating to the densities of animals when encountered (Zuur, Ieno, Walker, Saveliev, & Smith, 2009). These approaches helped combat statistical problems with zero-inflation and over-dispersion in the original data (Martin et al., 2005; Richards, 2008). The inclusion of a probability of encounters alongside animal densities provides two informative descriptors of species habitat use, discriminating between persistent presence of small groups and occasional presence of large groups. The hurdle approach also allowed scale-dependent processes to inform and influence SDM. For instance, biogeographical ranges are defined by presence–absence, and these usually coincide with environmental conditions influencing prey abundance (e.g. depth and temperature). By contrast, aggregations of animals within this range are defined by densities, and likely coincide with environmental conditions influencing prey availability (e.g. fronts and seabed roughness; Cox, Embling, Hosegood, Votier, & Ingram, 2018). Therefore, the presence–absence model should identify a biogeographical range, while the count model should identify aggregations of animals within this range.

Generalized Linear Models (GLM) and General Estimating Equations (GEE) (Koper & Manseau, 2009) using linear and quadratic terms were preferred over Generalized Additive Models (GAM; Wood, 2006). By misrepresenting the ecological niche of species, overfitting and underfitting model parameters represent serious issues in SDM (Elith & Leathwick, 2009). The complex relationships in GAM are susceptible to overfitting, whilst the simpler ones in GLM are vulnerable to underfitting (Derville, Torres, Iovan, & Garrigue, 2018). It was believed that heterogeneous and uneven coverage of survey data could cause overfitting in GAM. In particular, model parameters could be overly influenced by artificially enhanced counts in areas of intense coverage, a particularly large count in areas of low coverage, or anomalous counts during unusual environmental conditions. By contrast, it was considered the large amounts of survey data would reduce the likelihood of underfitting in GLM. More specifically, there should be sufficient information to identify the ecological niche of each species (Stockwell & Peterson, 2002). GEE were used to account for any spatial and temporal autocorrelation in the residuals of GLM. GEE-adjusted model parameters were based on correlations among surveys from the same supplier and month.

A binomial family with a logit link function was used for the presence–absence model, with the presence/absence of a species as the response variable. The surface area searched per cell (km^2) was included as a statistical offset to account for variations in effort among samples. For seabirds, where there were two measurements per cell, the surface area searched per cell represented the mean of that for animals on the sea surface and those in flight. Due to intense coverage in certain cells, the offset was log-transformed. This was on the assumption that the probability of encounters reaches a threshold when large surface areas have been covered, i.e. species have already been found if present. A Poisson family was used for the count model, with the square-root transformed density of animals as the response variable. Usually numbers of animals are used as a response variable, with a statistical offset used to account for variations in effort (Zuur et al., 2009). However, there was extreme overdispersion in the numbers of animals. A transformation was needed to combat extreme overdispersion, as negative binomial models cannot currently be applied to GEE-GLM. Unfortunately, transformations cannot be accommodated alongside a statistical offset. Using densities of animals and omitting the statistical offset accounted for variations in effort, while also allowing a transformation to be performed. For seabirds, using densities also eliminated the need to combine measurements of the surface area searched for animals on the sea surface and those in flight within the statistical offset. A square-root rather than log-transformation was chosen because densities of animals could be <1 (Zar, 2010). Aforementioned environmental conditions were the explanatory variables in binomial and Poisson models (Table 4). GEE-GLM were performed using the 'GEEPACK' package (Højsgaard, Halekoh, & Yan, 2006) in R.

In the presence–absence model, the optimal model was selected using forwards-model selection (Zuur et al., 2009) based on quasi-likelihood under the model independence criterion (QIC). This approach allowed variables to be included at an appropriate scale, starting with those believed to have the largest influence on distributions. Those describing different biomes (1,000+ km) (depth, annual temperature variance) and breeding aggregations (colony index) were introduced first; those describing different areas (100–1,000 km) within these biomes (annual temperature, regional temperature) were introduced second. In the count model, the optimal model was selected using multi-model selection using QIC (Burnham & Anderson, 2002). This was because seabed roughness and fronts operate at a similar scale, describing features in an area (10–100 km). Only plausible relationships showing proven associations between animals and environmental conditions were allowed (Table 4).

2.3.4 | Predictions

The production of distribution maps focused upon the exclusive economic zones (EEZs) of (north to south) Norway, UK, Ireland, Sweden, Denmark and Germany, The Netherlands, Belgium, Atlantic France and northwest Spain (2,148,000 km^2) covered by the FOAM AMM7 model domain (discussed above). Densities (animals per km^2) were predicted at monthly and 10 km resolution for each species, using the appropriate GEE-GLM. The probabilities of encountering animals were estimated using the binomial model; the densities of animals if encountered were estimated using the Poisson model. The final density estimations were a product of these two components (Barry & Welsh, 2002). Values of environmental conditions were constrained between 5% and 95% quantiles of the minimum and maximum values to avoid unrealistic estimations of densities in areas with extreme conditions, e.g. estuaries and fjords. More specifically, values at 0%–5% and 95%–100% quantiles were replaced by those at exactly 5%–95% quantiles, respectively. GEE-GLM uncertainty per month and cell was quantified using 5%–95% quantiles of predicted densities from 1,000 simulations of parameter

TABLE 4 Summary of the forward-selection process in the binomial and Poisson model

Model	Stage	Candidate variable	Ecological reasoning	Relationships not accepted
Binomial	1	Breeding colony[b] + Breeding cycle[b]	Seabirds aggregate around large breeding colonies in summer months	Negative relationships, as the probability of encounters should not increase further from large breeding colonies in summer months
	2	Depth*	Prey communities are associated with particular depths	U-shaped relationships with depth, as associations with both extreme deep and shallow water are unlikely
		Depth* + Annual temperature variance	Prey communities are associated with particular depths, but avoid habitats characterised with unstable water conditions	
		Land[c]	European Shags regularly roost on land to dry-out their wettable plumage	Negative relationships, as the probability of encounters should not increase further offshore
	3	Annual temperature[a]	Prey communities are associated with long-term temperature	U-shaped relationships with annual temperature, as associations with both extreme cold and warm water are unlikely
		Annual temperature[a] + Regional temperature	Prey communities are associated with long-term temperature, but have seasonal variations in abundance	
		Annual temperature[a] + Regional temperature*Depth	Prey communities are associated with long-term temperature, but have seasonal variations in abundance and/or movements between shallow and deep water	
		Annual temperature[a] + Regional temperature*Annual temperature	Prey communities are associated with long-term temperature, but have seasonal variations in abundance and/or movements between cool and warm areas	
		Annual temperature[a] + Regional temperature*Annual temperature variance	Prey communities are associated with long-term temperature, but have seasonal variations in abundances and/or movements between stable and instable areas	
Poisson	1	Seabed roughness	Areas of rough seabed create hydrodynamic processes that increase the availability of pelagic prey. Those of smooth seabeds accumulate sediment and increase the availability of demersal and benthic prey	None
		Fronts	The presence of fronts creates hydrodynamic processes that increase the availability of pelagic prey	Negative relationships, as it is unclear how the absence of fronts could enhance prey availability

Note: Quasilikelihood under the model independence criterion (QIC) was used to select the best option at each stage.

[a]Quadratic relationships.

[b]Relationships exclusive to seabirds.

[c]Relationships exclusive to European Shag.

*interaction term.

estimates. Simulated parameter estimates followed a normal distribution, with variance around the mean determined by the covariance matrix. Estimations of uncertainty were performed using the 'MVTNORM' package (Genz et al., 2017) in R.

Model performance was evaluated qualitatively using knowledge of species distributions in the study area, and quantitatively using area under the curve (AUC) and normalized root-mean-squared-error (NRMSE). AUC describes the ability of the binomial model to predict presences and absences in the original observations. NRMSE represents the mean difference between predicted and observed values in the Poisson model, standardized using the

range in the latter. Both produce indices with values between 0 and 1. AUC values approaching 1 and NRMSE approaching 0 represent better performance.

3 | RESULTS

3.1 | Collation

Detailed summaries of the survey data including coverage, data suppliers, platforms/transect methods, and numbers of sightings are

provided in the supporting information (Figures S1–S2, Tables S2–S4). 2,682,363 and 1,649,297 km of survey data were collated for cetaceans and seabirds, respectively. There was a notable contribution of non-government organizations (NGOs) within survey data (35%).

3.2 | Standardization

Tables 5 and 6 provides a summary of *esw* and *g*(0) estimations, respectively. The probability of detection up to the maximum *esw* (300 m for ESAS, 1 km for line-transects) generally increased with body size, being greatest in fin whales/sperm whales for cetaceans and northern gannets for seabirds. The probability of detection was generally larger in ESAS than line-transects. By contrast, the probability of detection showed no consistent differences between aircraft and vessels. However, substantial differences between aerial and vessel line-transects were present for fin whales and sperm whales. An influence of sea state and platform height was commonplace for cetaceans from line-transect surveys. Such an influence was less frequent for ESAS and seabirds. Estimates of *g*(0) from vessels were broadly similar among cetaceans, with the lowest values occurring in sperm whales and the highest in small dolphins (Atlantic white-sided, bottlenose, short-beaked common, striped and white-beaked dolphin). 1,790,375 and 1,143,587 km of survey data were available for cetacean and seabird SDM, respectively, following the removal of line-transects and ESAS in sea states greater than Beaufort scale 3.

3.3 | Species distribution models

3.3.1 | Environmental associations

Summaries of recorded densities used to quantify associations between each species and environmental conditions are provided in the supporting information (Figures S3–S4). Figures 1–3 show associations between species and environmental conditions.

Optimal temperatures and depths tended to be higher in cetaceans than seabirds. Relationships with annual temperature variance differed among species, although cetaceans generally showed stronger relationships than seabirds. All cetaceans and seabirds showed relationships with regional temperature. The ever-presence of interactions involving regional temperature indicated that seasonal movements across environmental gradients are commonplace. Movements across latitudes were the most prevalent seasonal movement, although movements across gradients in depth and habitat stability were frequent. Relationships with fronts and/or rough seabed's were frequent.

Seabird relationships with colony indices differed in strength, indicating variations in associations with large breeding colonies. Relationships with breeding season also differed in whether species were detected more in breeding or non-breeding seasons. The former presumably identifies migratory species moving into the

region. The latter probably identifies those abundant year-round, with overall numbers of animals decreasing in breeding seasons when populations are divided between marine and terrestrial areas.

3.3.2 | Predictions

Predicted distributions, uncertainty in predicted distributions, and differences in predicted distributions between months are provided in the supporting information (Appendix S1–S3). Predicted distributions for January and July are shown in Figures 4 and 5 to demonstrate variation between coolest and warmest months, respectively.

Qualitative assessment using prior knowledge indicated good model performance. Long-distance migrants (Procellariiformes and Mysticetes) moved into the region *en-masse* during summer (Evans, 2008; Snow & Perrins, 2004). Odontocetes believed to be abundant year-round (bottlenose dolphin, harbour porpoise, long-finned pilot whale, short-beaked common dolphin, sperm whale) persisted in the region, whereas transient odontocetes moved into the region during summer (Atlantic white-sided dolphin, killer whale, Risso's dolphin, striped dolphin, white-beaked dolphin; Reid et al., 2003). Seabirds considered to be abundant year-round (black-legged kittiwake, common guillemot, European shag, herring gull, razorbill) aggregated around colonies in summer, and dispersed across the region in winter (Kober et al., 2010; Stone et al., 1995). Those considered to as transient (Atlantic puffin, great skua, lesser black-backed gulls, northern fulmar, northern gannet) aggregated around colonies in summer, before moving outside the region in winter (Kober et al., 2010; Stone et al., 1995). Quantitative assessment also showed consistently good model performance. AUC values for binomial models were always greater than 0.75—exceeding 0.80 on 18/24 occasions and 0.90 on 10/24 occasions (Table 7). While NRMSE values for Poisson models varied more amongst species, differences between predicted and observed densities never exceeded 21% of the observed density range—being less than 10% on 20/24 occasions and 5% on 9/24 occasions (Table 7).

4 | DISCUSSION

This study developed approaches to produce distributional maps for 12 cetacean and 12 seabird species at 10 km and monthly resolution in the North-East Atlantic. This process was divided into three stages: collation of survey data, standardization of survey data, and SDM.

4.1 | Collation

This study provides the largest collation of its kind for cetaceans, exceeding previous ones from the Mediterranean (Mannocci et al., 2018), western Atlantic (Roberts et al., 2016) and the British EEZ (Paxton et al., 2016). As it includes and supplements the largest existing collation from the North-East Atlantic (Kober et al., 2010), it is

TABLE 5 Summary of esw calculations for cetaceans and seabirds: sample size (n), response type (hr = hazard rate, hn = half normal: Res), slope estimate for platform height (PL), slope estimate for sea state (SS), probability of detection up to the maximum *esw* (Pr), standard error in the probability of detection up to the maximum *esw* (Se) and coefficient of variation in probability of detection up to the maximum *esw* (CV)

Taxa	Species	Behaviour	ESAS vessel (300 m)							Line vessel (1 km)							Line aerial (1 km)						
			n	Res	PL	SS	Pr	Se	CV	n	Res	PL	SS	Pr	Se	CV	n	Res	PL	SS	Pr	Se	CV
Cetacean	AWSD, BND, SBCD, SD, WBD	On Water	2,206	hr	0.00	-0.65	0.45	0.05	0.11	7,625	hr	0.55	-0.47	0.14	0.00	0.03	2,140	hr	0.00	-0.16	0.21	0.00	0.02
	HP	On Water	2,544	hr	0.00	0.00	1.00	0.00	0.00	9,026	hr	0.30	-0.27	0.24	0.00	0.01	13,987	hr	-0.50	-0.05	0.20	0.00	0.01
	FW	On Water	55	hn	0.00	0.00	1.00	0.10	0.10	958	hn	0.64	0.00	0.89	0.03	0.04	102	hr	0.00	-0.24	0.44	0.03	0.06
	KW, LFPW, RD	On Water	274	hn	0.00	0.00	1.00	0.05	0.05	673	hr	0.38	-0.85	0.38	0.04	0.10	227	hr	0.00	-0.16	0.33	0.02	0.06
	MW	On Water	294	hn	0.00	0.00	1.00	0.05	0.05	1,463	hr	0.20	-0.20	0.31	0.02	0.05	157	hr	0.00	0.00	0.27	0.02	0.08
	SPW	On Water	64	hn	0.00	0.00	1.00	0.08	0.08	166	hn	0.00	0.00	0.96	0.09	0.09	27	hn	0.00	0.00	0.49	0.08	0.16
Seabird	ESP	Flight	—	—	—	—	—	—	—	129	hr	0.00	0.00	0.12	0.01	0.12	46	hn	0.00	0.00	0.16	0.03	0.18
		On Water	745	hn	2.98	0.00	0.97	0.02	0.02	15	hn	1.86	0.00	0.22	0.07	0.30	1	hr	0.00	0.00	0.10	0.00	0.01
	FUL	Flight	—	—	—	—	—	—	—	623	hr	0.00	0.00	0.16	0.01	0.06	2,233	hr	0.00	0.00	0.28	0.00	0.01
		On Water	32,982	hn	6.70	-0.25	0.99	0.00	0.00	130	hr	0.00	0.00	0.20	0.02	0.10	636	hr	0.00	0.00	0.25	0.01	0.02
	GAN	Flight	—	—	—	—	—	—	—	5,919	hr	0.45	0.00	0.33	0.01	0.02	8,598	hr	0.00	-0.26	0.42	0.00	0.01
		On Water	18,064	hr	0.00	0.00	1.00	0.00	0.00	1989	hr	0.21	0.00	0.37	0.01	0.03	3,433	hr	0.00	-0.16	0.41	0.01	0.02
	GIL, PUF, RAZ	Flight	—	—	—	—	—	—	—	461	hr	0.00	0.00	0.17	0.01	0.07	2,677	hr	0.00	-0.04	0.27	0.00	0.01
		On Water	125,230	hr	0.95	-0.92	0.84	0.00	0.00	1,128	hr	0.00	0.00	0.23	0.01	0.03	45,997	hr	0.00	0.00	0.26	0.00	0.00
	GRK	Flight	—	—	—	—	—	—	—	615	hr	0.47	0.00	0.29	0.01	0.05	77	hr	0.00	0.00	0.26	0.02	0.08
		On Water	1,346	hr	0.00	0.00	1.00	0.00	0.00	118	hr	0.72	-0.26	0.39	0.03	0.08	12	hn	0.00	0.00	0.22	0.06	0.26
	HEG, LBB	Flight	—	—	—	—	—	—	—	2,664	hr	0.00	0.00	0.20	0.00	0.02	5,249	hr	0.00	0.00	0.27	0.00	0.01
		On Water	15,285	hr	0.00	0.00	1.00	0.00	0.00	562	hr	0.00	0.00	0.30	0.01	0.05	1,028	hr	0.00	0.00	0.27	0.01	0.02
	KIT	Flight	—	—	—	—	—	—	—	248	hr	0.00	-0.58	0.19	0.01	0.08	10,648	hr	0.00	-0.02	0.27	0.00	0.01
		On Water	12,047	hr	0.00	-0.47	0.74	0.01	0.02	47	hn	0.00	0.00	0.25	0.02	0.09	2,181	hr	0.00	0.00	0.25	0.00	0.01
	MSH	Flight	—	—	—	—	—	—	—	140	hr	0.63	0.00	0.21	0.02	0.10	2,220	hr	0.00	0.00	0.27	0.00	0.01
		On Water	2,603	hn	2.01	-0.96	0.97	0.01	0.01	8	hr	0.00	0.00	0.12	0.06	0.53	596	hr	0.00	0.00	0.29	0.01	0.03
	SHG	Flight	—	—	—	—	—	—	—	78	hn	0.00	0.00	0.28	0.03	0.09	79	hr	0.00	0.00	0.27	0.02	0.09
		On Water	919	hr	0.00	0.00	1.00	0.00	0.00	20	hn	0.00	0.00	0.34	0.06	0.17	440	hr	0.00	0.00	0.28	0.01	0.04

Note: Esw was not calculated for flying seabirds from ESAS vessels that always use a strip-transect. Species codes are outlined in Table 1. Explanatory variables are described in Table 2.

TABLE 6 Summary of g(0) calculations for cetaceans from vessel surveys: sample size (*n*), slope estimate of platform height (PL), slope estimate of sea state (SS), estimations of g(0), standard error in g(0) (Se) and coefficient of variation in g(0) (CV). g(0) estimations for cetaceans from aerial surveys were sourced from existing studies using biologging techniques. g(0) for vessel surveys accounts for availability, perception and response bias; those for aerial surveys accounts for availability bias only

	Vessel						Aerial	
Species	*n*	**PL**	**SS**	**g(0)**	**Se**	**CV**	**g(0)**	**Source**
AWSD, BND, SBCD, SD, WSD	2024	0.00	0.00	0.58	0.09	0.16	0.82	Rasmussen et al. (2013)
HP	5,122	0.00	0.00	0.31	0.04	0.11	0.19	Hansen et al. (2018)
FW	66	0.00	0.00	0.53	0.25	0.47	0.19	Hansen et al. (2018)
KW, LFPW, RD	164	0.00	0.00	0.49	0.15	0.30	0.76	Alves et al. (2013)
MW	610	−0.33	0.00	0.40	0.13	0.33	0.16	Hansen et al. (2018)
SPW	32	0.00	0.00	0.25	0.20	0.80	0.17	Watwood et al. (2006)

Note: Species codes are outlined in Table 1. Explanatory variables are described in Table 2.

also the largest of its kind for seabirds. A particular characteristic of this collation is the sizeable contribution from NGOs. These organizations are independently funded, drawing heavily from the voluntary sector. As a consequence, they are usually conducted on vessels of opportunity (e.g. continental and regional ferries) and/or on those chartered from local commercial operators (Evans & Hammond, 2004). This study demonstrates the invaluable resource provided by NGOs. This importance is most evident in the detection of seasonal movements, made possible through intensive coverage of particular areas across different months.

4.2 | Standardization

While the approaches used to standardize surveys are not novel, this study is one of few applications of these approaches (Paxton et al., 2016). The considerable variations in *esw* and g(0) indicate that differences in surface area searched occur among surveys, and supports the use of this metric to standardize diverse survey data. However, the absence of g(0) for seabirds could have limited the comparability of vessel and aerial surveys. In particular, scavenging species (Laridae, northern gannets and northern
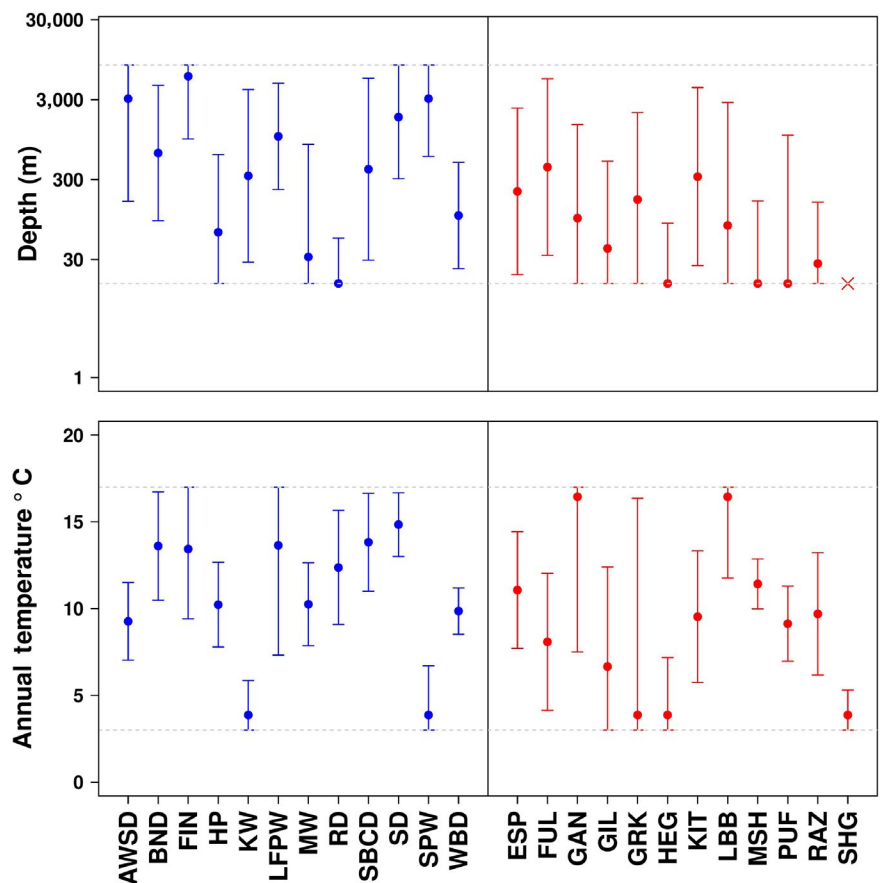
FIGURE 1 Summary of quadratic relationships between species and annual temperature/depth in the North-East Atlantic, as quantified using a binomial General Estimating Equations-Generalized Linear Models. Points indicate values where the probability of encounters were highest, whereas lines indicate values for 25% and 75% quantiles around the highest probabilities. The dashed lines indicate the minimum and maximum values of annual temperature and depth in the study area. Cetaceans are shown in blue, and seabirds are shown in red. Crosses indicate when a relationship was not identified. Species codes are described in Table 1
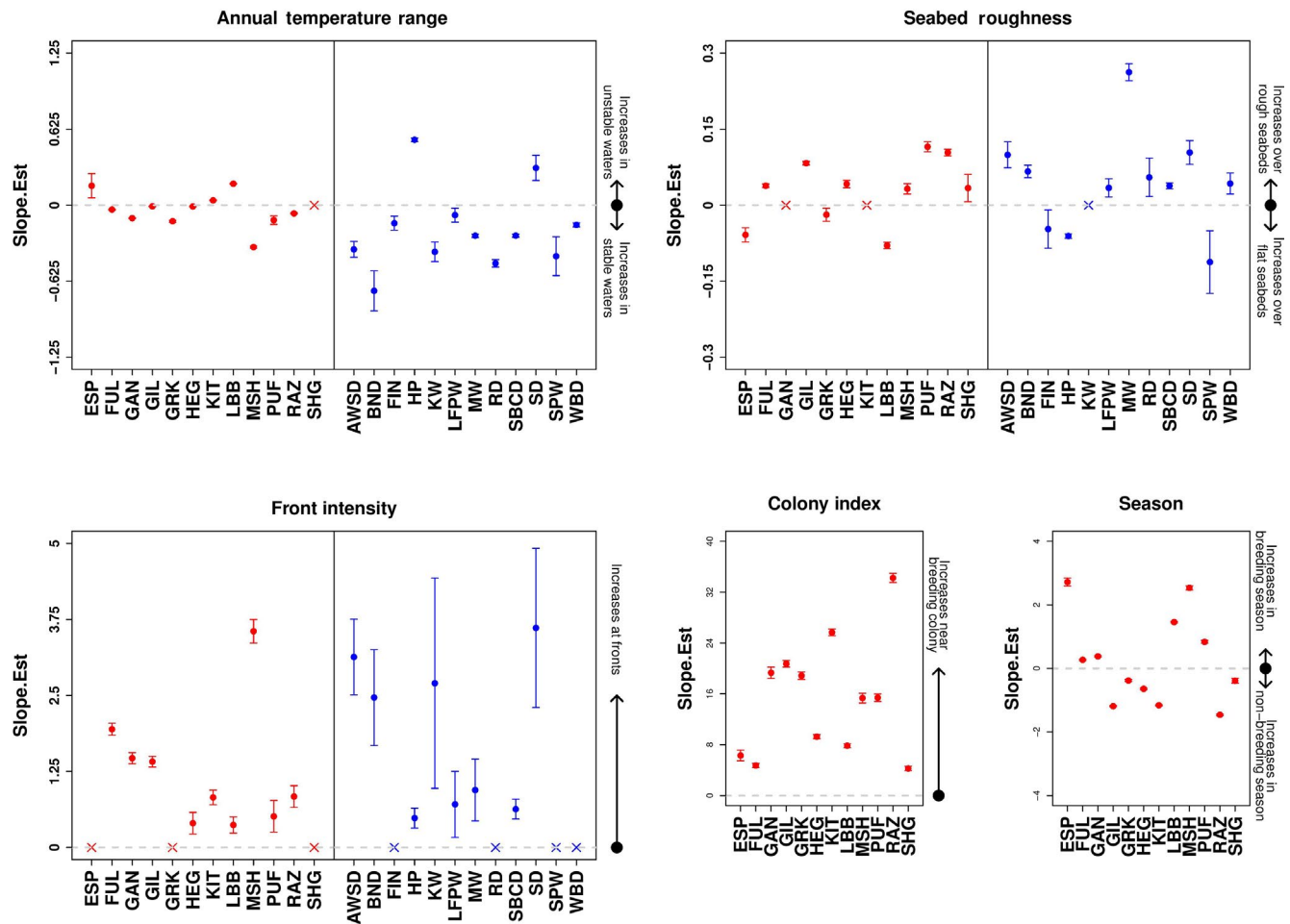
FIGURE 2 Summary of linear relationships between species and environmental conditions in the North-East Atlantic, as quantified using a binomial (annual temperature range, colony index, season) or Poisson (seabed roughness, front intensity) General Estimating Equations-Generalized Linear Models. Points indicate slope estimates, whereas lines indicate standard errors around this estimate. The dashed line indicates a slope estimate of 0. Crosses indicate when a relationship was not identified. Information onand environmental conditions are provided in the environmental conditions is in Table 4. Cetaceans are shown in blue, and seabirds are shown in red. Species codes are described in Table 1

fulmars) will readily approach vessels but not aircraft, resulting in response bias in the former but not the latter. The calculation of g(0) requires the performance of double-platform transects. Unfortunately, these transects are rarely implemented for seabirds from vessels. This absence is possibly because attraction bias is rarely considered and/or availability bias is assumed to be negligible as animals are mainly in flight or on the sea surface (Ronconi & Burger, 2009). Therefore, the standardization of seabird surveys could be improved.

4.3 | Species distribution models

The study aimed to quantify basin and monthly-scale distributions of species, whilst overcoming problems with heterogeneous and uneven coverage. This led to the development of models that differed from conventional SDM approaches. Firstly, GEE-GLM rather than GAM approaches were chosen to reduce

overfitting, producing distribution maps that illustrated a species range rather than areas/times of intense effort. Hurdle-model approaches were also chosen to combine information on the probabilities of encounters and the animals densities if encountered (Zuur et al., 2009), preventing occasional encounters with large groups having a greater influence on models parameters than persistent encounters with small groups. It appears that these aims were met; outputs did not give strong prominence to particular areas, did not contain extreme outliers, and showed similarities to sightings Atlases (Reid et al., 2003; Stone et al., 1995). Secondly, interactions between annual and monthly averaged temperatures rather than concurrent temperatures were used as explanatory variables, covering a broader range of seasonal movements. In some cases, it appears that these aims were also met; outputs showed seasonal movements that would not have been detected using concurrent temperatures. For instance, that of long-finned pilot whale and sperm whale into deeper waters during summer months, and of harbour porpoise into the innermost North
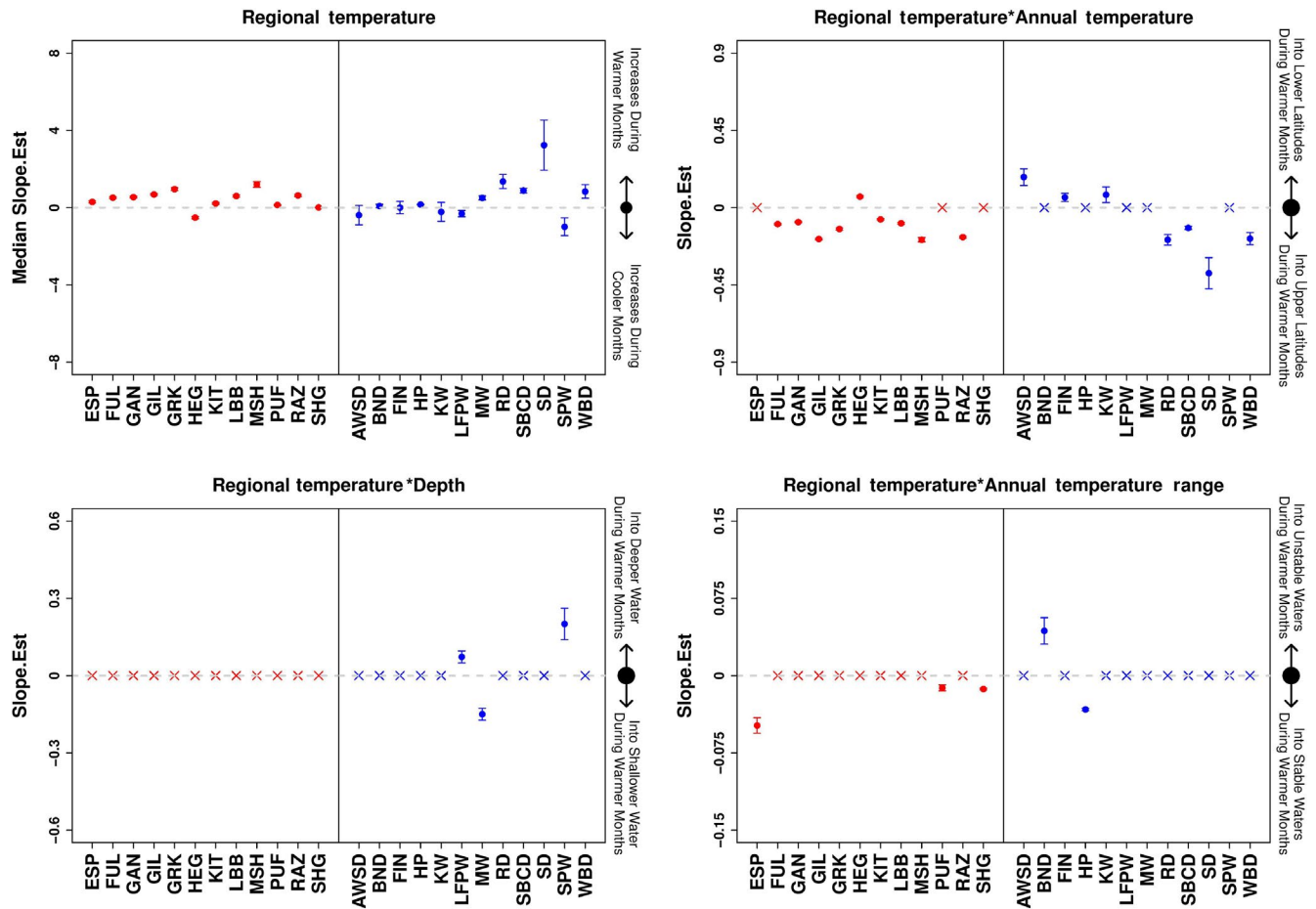
FIGURE 3 Summary of linear interactive relationships between species and environmental conditions in the North-East Atlantic, as quantified with a binomial General Estimating Equations-Generalized Linear Models. Points indicate slope estimates, whereas lines indicate standard errors around this estimate. Crosses indicate where a relationship was not identified. Information on environmental conditions is in Table 4. Cetaceans are shown in blue, and seabirds are shown in red. Species codes are described in Table 1

Sea during winter months. Assessment showed that model performance was not compromised by using non-conventional approaches. This emphasizes the usefulness of developing bespoke methods tailored to the data properties and the study aims (Derville et al., 2018).

4.4 | Limitations

The distribution maps need careful interpretation. Firstly, small and isolated sub-populations would have little influence on models. Examples include white-beaked dolphins in south-west England (Brereton, Lewis, & MacLeod, 2012) and Risso's dolphins in North Wales/Isle of Man (Baines & Evans, 2012). Second, there have been substantive changes in populations across the study period. For instance, the core-distribution of harbour porpoise has moved from the northern to the southern North Sea in recent years (Hammond et al., 2013), while seabird numbers have declined in the northern North Sea (SNH, 2012). Thirdly, despite seasonal movements being detected, seasonal increases and decreases in densities without notable changes in distribution were more commonplace. This general

absence could indicate constraints imposed by the SDM setup, and complicated or inconsistent seasonal movements amongst years. Finally, uncertainty on the sizes of seabird colonies (Mitchell, Newton, Ratcliffe, & Dunn, 2004) could lead to SDM-induced biases where numbers of breeding animals have been misrepresented. Because of these caveats, outputs should not be used as a representation of absolute densities and fine-scale distributions at the present time. Instead, it is recommended that outputs be used as a general illustration of relative densities and broad-scale distribution over several decades.

4.5 | Applications

This study provides the most comprehensive cetacean and seabird distribution maps at basin and seasonal-scales in Europe (Kober et al., 2010; Paxton et al., 2016). The quantity and extent of survey data in the collation should provide a good representation of distributional patterns in the study area. The ecologically informed SDM setup also enables patterns to be supported with realistic environmental associations based on empirical
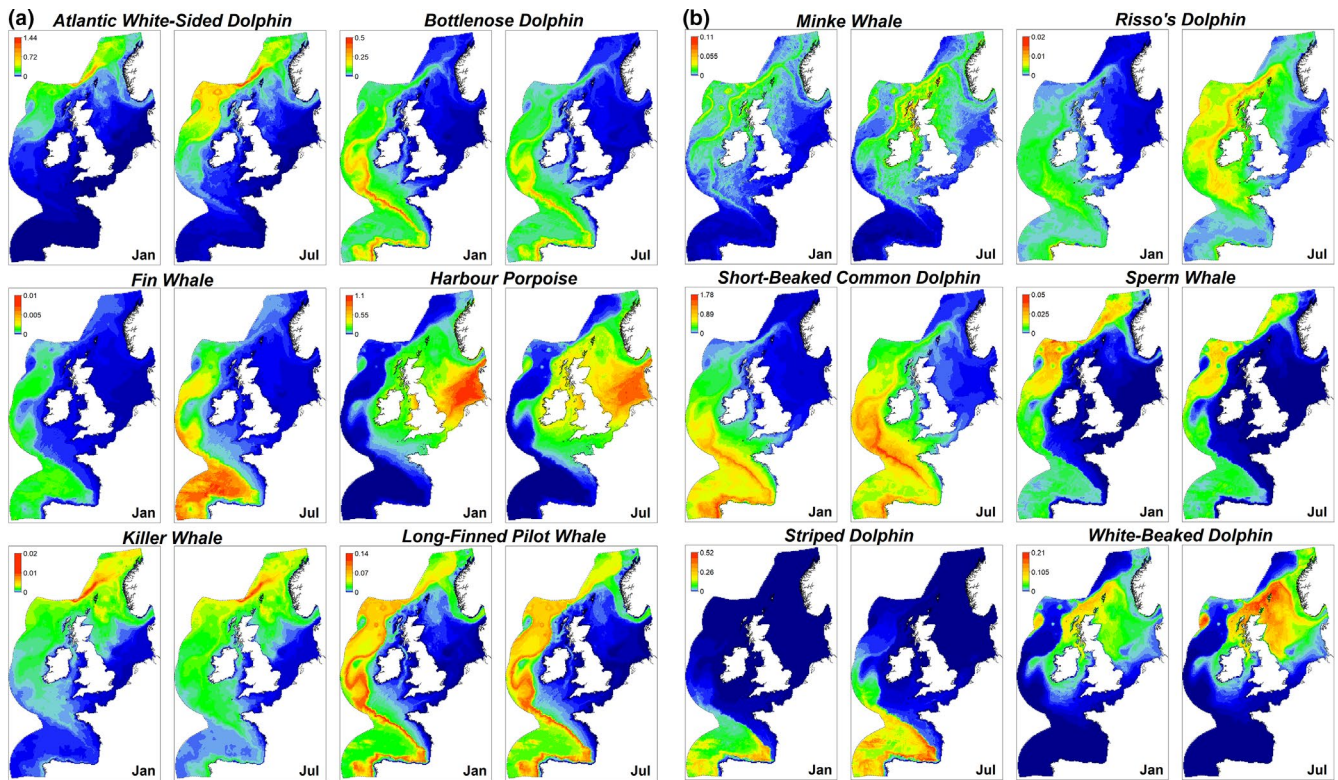
FIGURE 4 (a, b) Spatial variation in predicted densities (animals per km^2) of cetacean species in January and July in the North-East Atlantic. Values are provided at 10 km resolution. A different colour gradient is used for each species. Bottlenose dolphin in (a) represent the offshore ecotype
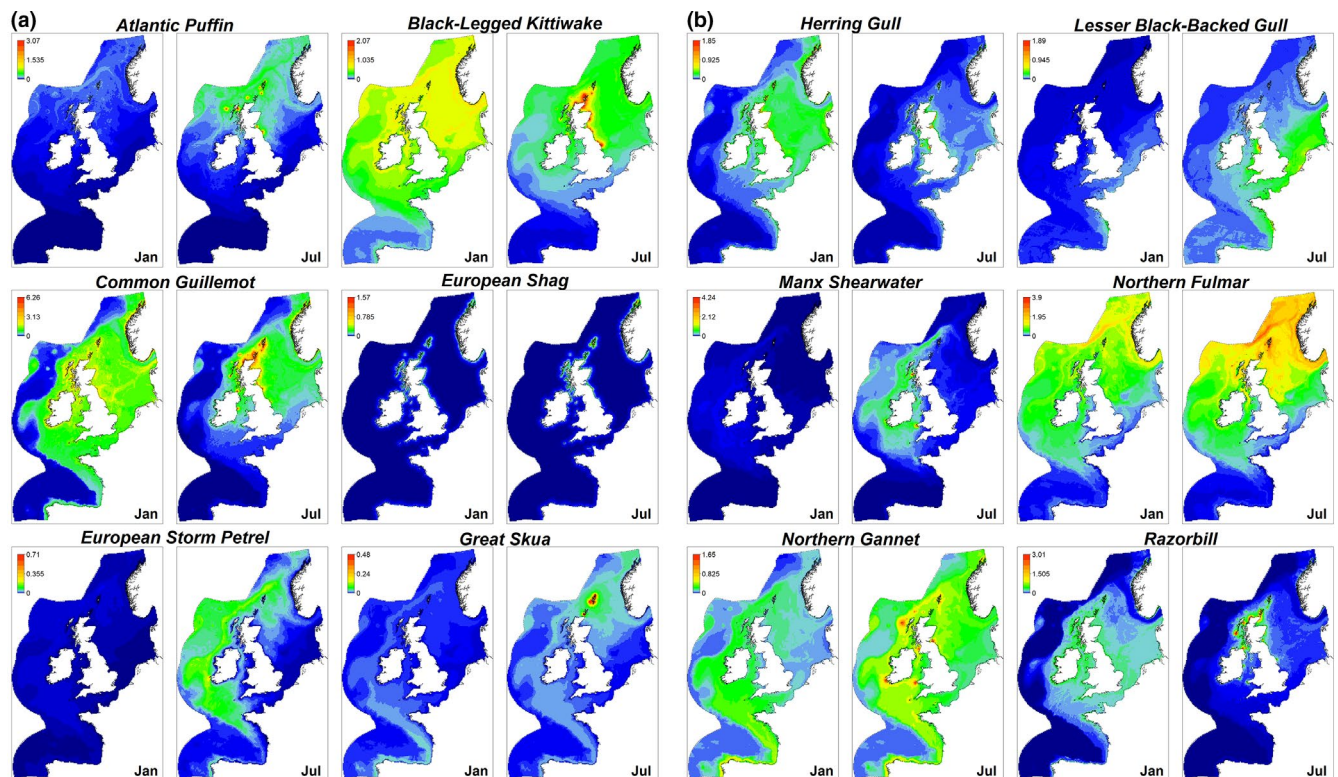

FIGURE 5 (a, b) Spatial variation in predicted densities (animals per km^2) of seabird species in January and July in the North-East Atlantic. Values are provided at 10 km resolution. A different colour gradient is used for each species

TABLE 7 Quantitative evaluation of presence-absence and density GEE-GLM predictions using area under the curve (AUC) and normalised root mean squared error (NRMSE), respectively

Taxa	Species	AUC	NRMSE
Cetacean	Atlantic White-Sided Dolphin	0.92	0.07
	Bottlenose Dolphin	0.91	0.09
	Fin Whale	0.96	0.17
	Harbour Porpoise	0.79	0.05
	Killer Whale	0.86	0.14
	Long-Finned Pilot Whale	0.93	0.04
	Minke Whale	0.79	0.09
	Risso's Dolphin	0.85	0.14
	Short-Beaked Common Dolphin	0.87	0.05
	Sperm Whale	0.97	0.21
	Striped Dolphin	0.98	0.07
	White-Beaked Dolphin	0.85	0.07
Seabird	Atlantic Puffin	0.91	0.05
	Black-Legged Kittiwake	0.78	0.03
	Common Guillemot	0.81	0.03
	European Shag	0.94	0.08
	European Storm Petrel	0.93	0.08
	Great Skua	0.83	0.08
	Herring Gull	0.79	0.03
	Lesser Black-Backed Gull	0.76	0.03
	Manx Shearwater	0.91	0.04
	Northern Fulmar	0.85	0.03
	Northern Gannet	0.77	0.02
	Razorbill	0.82	0.03

evidence; for example, the presence of scale-dependent associations between top-predators and environmental conditions (Cox et al., 2018). While some caution is needed, these distribution maps have widespread and immediate applications. For instance, combining distribution maps of vulnerable species and anthropogenic activities could identify when and where interactions are likely to occur, aiding the environmentally-responsible use of marine resources (Croxall et al., 2012; Evans & Anderwald, 2016). Distribution maps could also be used to identify important areas in need of protection (Evans, 2018; Lascelles, Langham, Ronconi, & Reid, 2012). This study demonstrates how analysis of existing and diverse data can meet conservation and marine management needs.

AUTHORS' CONTRIBUTIONS

J.J.W. wrote the manuscript, collated/standardized survey data, and developed SDM. P.G.H.E. and J.G.H. helped develop these approaches. The remaining co-authors contributed survey data, and revised the manuscript. All authors gave final approval for publication.

DATA AVAILABILITY STATEMENT

Distribution maps are available via the Dryad Digital Repository https://doi.org/10.5061/dryad.mw6m905sz (Waggitt et al., 2019). Any requests for survey data should be addressed to their owners. Contact details of the owners are provided in the supporting information (Table S5). In future, some survey data may become open access. Please contact PGHE (peter.evans@bangor.ac.uk) for further details.

ORCID

James J. Waggitt https://orcid.org/0000-0003-3172-7096
Peter G. H. Evans https://orcid.org/0000-0001-8139-9145
Ruben Christiaan Fijn https://orcid.org/0000-0002-6233-2954
Stefan Garthe https://orcid.org/0000-0003-0548-9346
Anita Gilles https://orcid.org/0000-0001-7234-8645
Mark Jessopp https://orcid.org/0000-0002-2692-3730
Ailbhe S. Kavanagh https://orcid.org/0000-0003-4671-3076
Sarah L. Perry https://orcid.org/0000-0001-8035-9600
Graham J. Pierce https://orcid.org/0000-0002-4744-4501
Vincent Ridoux https://orcid.org/0000-0002-9779-9319
Kevin P. Robinson https://orcid.org/0000-0001-8085-9752
M. Begoña Santos https://orcid.org/0000-0002-5759-1695
Camilo Saavedra https://orcid.org/0000-0003-1032-647X
Henrik Skov https://orcid.org/0000-0001-8835-3166
Eric W. M. Stienen https://orcid.org/0000-0002-4315-0936
Signe Sveegaard https://orcid.org/0000-0002-8893-9842
Paul Thompson https://orcid.org/0000-0001-6195-3284
Jan Geert Hiddink https://orcid.org/0000-0001-7114-830X

REFERENCES

Alves, F., Dinas, A., Ribero, C., Nicolau, C., Kaufman, M., Fortuna, C. M., & Freitas, L. (2013). Daytime dive characteristics from six short-finned pilot whales Globicephala macrorhynchus off Madeira Island. *Arquipelago - Life and Marine Sciences*, *8*, 1–13.

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

Baines, M. E., & Evans, P. G. H. (2012). *Atlas of the marine mammals of Wales. CCW Monitoring Report No. 68* (2nd ed.).

Barry, S. C., & Welsh, A. H. (2002). Generalized additive modelling and zero inflated count data. *Ecological Modelling*, *157*(2), 179–188. https://doi.org/10.1016/S0304-3800(02)00194-1

Boivin, N. L., Zeder, M. A., Fuller, D. Q., Crowther, A., Larson, G., Erlandson, J. M., … Petraglia, M. D. (2016). Ecological consequences of human niche construction: Examining long-term anthropogenic shaping of global species distributions. *Proceedings of the National*

Academy of Sciences, *113*(23), 6388–6396. https://doi.org/10.1073/pnas.1525200113

Borchers, D. L., Zucchini, W., Heide-Jørgensen, M. P., Cañadas, A., & Langrock, R. (2013). Using hidden markov models to deal with availability bias on line transect surveys. *Biometrics*, *69*(3), 703–713. https://doi.org/10.1111/biom.12049

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. https://doi.org/10.1016/j.ocecoaman.2016.09.004

Brereton, T. M., Lewis, K., & MacLeod, C. M. (2012). *Surveys for cetaceans, seabirds and other large marine animals in the English Channel*. Marinelife: Charm III Final Project Report.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., & Thomas, L. (2001). *Introduction to distance sampling*. Oxford, UK: Oxford University Press.

Buckland, S. T., Burt, M. L., Rexstad, E. A., Mellor, M., Williams, A. E., & Woodward, R. (2012). Aerial surveys of seabirds: The advent of digital methods. *Journal of Applied Ecology*, *49*(4), 960–967. https://doi.org/10.1111/j.1365-2664.2012.02150.x

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference* (Vol. 2nd). New York, UK: Springer.

Burt, M. L., Borchers, D. L., Jenkins, K. J., & Marques, T. A. (2014). Using mark–recapture distance sampling methods on line transect surveys. *Methods in Ecology and Evolution*, *5*(11), 1180–1191. https://doi.org/10.1111/2041-210X.12294

Camphuysen, C. J., Fox, A. D., Leopold, M. F., & Petersen, I. K. (2004). Towards standardised seabirds at sea census techniques in connection with environmental impact assessments for offshore wind farms in the U.K. Report by Royal Netherlands Institute for Sea Research and the Danish National Environmental Research Institute.

Cox, S. L., Embling, C. B., Hosegood, P. J., Votier, S. C., & 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. https://doi.org/10.1016/j.ecss.2018.06.022

Croxall, J. P., Butchart, S. H. M., Lascelles, B. G., Stattersfield, A. J., Sullivan, B., Symes, A., & Taylor, P. (2012). Seabird conservation status, threats and priority actions: A global assessment. *Bird Conservation International*, *22*, 1–34. https://doi.org/10.1017/S0959270912000020

Derville, S., Torres, L. G., Iovan, C., & Garrigue, C. (2018). Finding the right fit: Comparative cetacean distribution models using multiple data sources and statistical approaches. *Diversity and Distributions*, *24*(11), 1657–1673. https://doi.org/10.1111/ddi.12782

Doughty, C. E., Roman, J., Faurby, S., Wolf, A., Haque, A., Bakker, E. S., … Svenning, J.-C. (2016). Global nutrient transport in a world of giants. *Proceedings of the National Academy of Sciences*, *113*(4), 868–873. https://doi.org/10.1073/pnas.1502549112

Elith, J., & Leathwick, J. R. (2009). Species distribution models: Ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Evans, P. G. H. (2008). Whales, porpoises and dolphins. Order Cetacea. In S. Harris, & D. W. Yalden (Eds.), *Mammals of the British Isles* (4th ed., pp. 655–779). Southampton: The Mammal Society.

Evans, P. G. H. (2018). Marine protected areas and marine spatial planning for the benefit of marine mammals. *Journal of the Marine Biological Association of the United Kingdom*, *98*(5), 973–976. https://doi.org/10.1017/S0025315418000334

Evans, P. G. H., & Anderwald, P. (2016). Addressing human pressures upon marine mammals: A European and global perspective. *Journal of the Marine Biological Association of the United Kingdom*, *96*, 779–781. https://doi.org/10.1017/S0025315416000539

Evans, P. G. H., & Hammond, P. S. (2004). Monitoring cetaceans in European waters. *Mammal Review*, *34*(1–2), 131–156. https://doi.org/10.1046/j.0305-1838.2003.00027.x

Gaston, A. J. (2004). *Seabirds: A natural history*. New Haven, USA: Yale University Press.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2017). *1.0-6, mvtnorm: Multivariate normal and t distributions. R package version*.

Gilles, A., Viquerat, S., Becker, E. A., Forney, K. A., Geelhoed, S. C. V., Haelters, J., … Aarts, G. (2016). Seasonal habitat-based density models for a marine top predator, the harbor porpoise, in a dynamic environment. *Ecosphere*, *7*(6), e01367. https://doi.org/10.1002/ecs2.1367

Grecian, W. J., Witt, M. J., Attrill, M. J., Bearhop, S., Godley, B. J., Grémillet, D., … Votier, S. C. (2012). A novel projection technique to identify important at-sea areas for seabird conservation: An example using Northern gannets breeding in the North East Atlantic. *Biological Conservation*, *156*, 43–52. https://doi.org/10.1016/j.biocon.2011.12.010

Halpern, B. S., Frazier, M., Potapenko, J., Casey, K. S., Koenig, K., Longo, C., … Walbridge, S. (2015). Spatial and temporal changes in cumulative human impacts on the world's ocean. *Nature Communications*, *6*, 7615. https://doi.org/10.1038/ncomms8615

Halpern, B. S., Walbridge, S., Selkoe, K. A., Kappel, C. V., Micheli, F., Agrosa, C., … Watson, R. (2008). A global map of human impact on marine ecosystems. *Science*, *319*(5865), 948–952. https://doi.org/10.1126/science.1149345

Hammond, P. S. (2010). Estimating the abundance of marine mammals. In I. L. Boyd, W. D. Bowen, & S. J. Iverson (Eds.), *Marine mammal ecology and conservation. A handbook of techniques* (pp. 42–67). Oxford: Oxford University Press.

Hammond, P. S., Berggren, P., Benke, H., Borchers, D. L., Collet, A., Heide-Jørgensen, M. P., … Øien, N. (2002). Abundance of harbour porpoise and other cetaceans in the North Sea and adjacent waters. *Journal of Applied Ecology*, *39*(2), 361–376. https://doi.org/10.1046/j.1365-2664.2002.00713.x

Hammond, P. S., Macleod, K., Berggren, P., Borchers, D. L., Burt, L., Cañadas, A., … Vázquez, J. A. (2013). Cetacean abundance and distribution in European Atlantic shelf waters to inform conservation and management. *Biological Conservation*, *164*, 107–122. https://doi.org/10.1016/j.biocon.2013.04.010

Hansen, R. G., Boye, T. K., Larsen, R. S., Nielsen, N. H., Tervo, O., Nielsen, R. D., … Heide-Jørgensen, M. P. (2018). Abundance of whales in West and East Greenland in summer 2015. *BioRxiv*, 391680.

Hays, G. C., & Scott, R. (2013). Global patterns for upper ceilings on migration distance in sea turtles and comparisons with fish, birds and mammals. *Functional Ecology*, *27*(3), 748–756. https://doi.org/10.1111/1365-2435.12073

Higham, J., & Lück, M. (2007). *Marine Wildlife and Tourism Management: Insights from the Natural and Social Sciences*. Wallingford: CABI Publishing.

Hijmans, R. J. (2013). *Raster: Geographic data analysis and modelling. R package version 2.1-66*. Retrieved from http://CRAN.R-project.org/package=raster

Hoegh-Guldberg, O., & Bruno, J. F. (2010). The impact of climate change on the World's marine ecosystems. *Science*, *328*(5985), 1523–1528. https://doi.org/10.1126/science.1189930

Hoelzel, A. R., Potter, C. W., & Best, P. B. (1998). Genetic differentiation between parapatric 'nearshore' and 'offshore' populations of the bottlenose dolphin. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1402), 1177–1183. https://doi.org/10.1098/rspb.1998.0416

Højsgaard, S., Halekoh, U., & Yan, J. (2006). The R Package geepack for generalized estimating equations. *Journal of Statistical Software*, *15*(2), 1–11.

Hunt, G. L., & McKinnell, S. (2006). Interplay between top-down, bottom-up, and wasp-waist control in marine ecosystems. *Progress in Oceanography*, *68*(2), 115–124. https://doi.org/10.1016/j.pocean.2006.02.008

Kober, K., Webb, A., Win, I., Lewis, M., O'Brien, S., Wilson, L. J., & Reid, J. B. (2010). *An analysis of the numbers and distribution of seabirds within the British Fishery Limit aimed at identifying areas that qualify as possible marine SPAs. Joint Nature Conservation Committee Report 431.*

Koper, N., & Manseau, M. (2009). Generalized estimating equations and generalized linear mixed-effects models for modelling resource selection. *Journal of Applied Ecology*, *46*(3), 590–599. https://doi.org/10.1111/j.1365-2664.2009.01642.x

Kubetzki, U., & Garthe, S. (2003). Distribution, diet and habitat selection by four sympatrically breeding gull species in the south-eastern North Sea. *Marine Biology*, *143*(1), 199–207. https://doi.org/10.1007/s00227-003-1036-5

Lascelles, B. G., Langham, G. M., Ronconi, R. A., & Reid, J. B. (2012). From hotspots to site protection: Identifying Marine Protected Areas for seabirds around the globe. *Biological Conservation*, *156*, 5–14. https://doi.org/10.1016/j.biocon.2011.12.008

Louis, M., Viricel, A., Lucas, T., Peltier, H., Alfonsi, E., Berrow, S., … Simon-Bouhet, B. (2014). Habitat-driven population structure of bottlenose dolphins, Tursiops truncatus, in the North-East Atlantic. *Molecular Ecology*, *23*(4), 857–874. https://doi.org/10.1111/mec.12653

Mannocci, L., Roberts, J. J., Halpin, P. N., Authier, M., Boisseau, O., Bradai, M. N., … Vella, J. (2018). Assessing cetacean surveys throughout the Mediterranean Sea: A gap analysis in environmental space. *Scientific Reports*, *8*(1), 3126. https://doi.org/10.1038/s41598-018-19842-9

Martin, T. G., Wintle, B. A., Rhodes, J. R., Kuhnert, P. M., Field, S. A., Low-Choy, S. J., … Possingham, H. P. (2005). Zero tolerance ecology: Improving ecological inference by modelling the source of zero observations. *Ecology Letters*, *8*(11), 1235–1246. https://doi.org/10.1111/j.1461-0248.2005.00826.x

Mitchell, P. I., Newton, S. F., Ratcliffe, N., & Dunn, T. E. (2004). *Seabird populations of Britain and Ireland*. London, UK: Poyser.

Paxton, C. G. M., Scott-Hayward, L., Mackenzie, M., Rexstad, M., & Thomas, L. (2016). *Revised phase III data analysis of joint cetacean protocol data resource. JNCC Report No517.*

R Development Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rasmussen, M. H., Akamatsu, T., Teilmann, J., Vikingsson, G., & Miller, L. A. (2013). Biosonar, diving and movements of two tagged white-beaked dolphin in Icelandic waters. *Deep Sea Research Part II: Topical Studies in Oceanography*, *88–89*, 97–105. https://doi.org/10.1016/j.dsr2.2012.07.011

Reid, J. B., Evans, P. G. H., & Northridge, S. P. (2003). *Atlas of cetacean distribution in north-west European waters*.

Richards, S. A. (2008). Dealing with overdispersed count data in applied ecology. *Journal of Applied Ecology*, *45*(1), 218–227. https://doi.org/10.1111/j.1365-2664.2007.01377.x

Roberts, J. J., Best, B. D., Mannocci, L., Fujioka, E. I., Halpin, P. N., Palka, D. L., … Lockhart, G. G. (2016). Habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico. *Scientific Reports*, *6*, 22615. https://doi.org/10.1038/srep22615

Ronconi, R., & Burger, A. E. (2009). Estimating seabird densities from vessel transects: Distance sampling and implications for strip transects. *Aquatic Biology*, *4*(3), 297–309. https://doi.org/10.3354/ab00112

SNH. (2012). *Trend note: Seabirds in Scotland. Report by Scottish Natural Heritage*.

Snow, D., & Perrins, C. M. (2004). *The Birds of the Western Palaearctic (Concise Edition)*. Oxford: Oxford University Press.

Stockwell, D. R. B., & Peterson, A. T. (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, *148*(1), 1–13. https://doi.org/10.1016/S0304-3800(01)00388-X

Stone, C. J., Webb, A., Barton, C., Ratcliffe, N., Reed, T. C., Tasker, M. L., … Pienkowski, M. W. (1995). *An atlas of seabird distribution in North-West European waters*. Peterborough, UK: Joint Nature Conservation Committee.

Thaxter, C. B., Lascelles, B., Sugar, K., Cook, A. S. C. P., Roos, S., Bolton, M., … Burton, N. H. K. (2012). Seabird foraging ranges as a preliminary tool for identifying candidate Marine Protected Areas. *Biological Conservation*, *156*, 53–61. https://doi.org/10.1016/j.biocon.2011.12.009

Thaxter, C. B., Wanless, S., Daunt, F., Harris, M. P., Benvenuti, S., Watanuki, Y., … Hamer, K. C. (2010). Influence of wing loading on the trade-off between pursuit-diving and flight in common guillemots and razorbills. *Journal of Experimental Biology*, *213*, 1018–1025. https://doi.org/10.1242/jeb.037390

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

Waggitt, J. J., Evans, P. G. H., Andrade, J., Banks, A. N., Boisseau, O., Bolton, M., … Hiddink, J. G. (2019). Data from: Distribution maps of cetacean and seabird populations in the North-East Atlantic. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.mw6m905sz

Wanless, S., Corfield, T., Harris, M. P., Buckland, S. T., & Morris, J. A. (1993). Diving behaviour of the shag *Phalacrocorax aristotelis* (Aves: Pelecaniformes) in relation to water depth and prey size. *Journal of Zoology*, *231*(1), 11–25. https://doi.org/10.1111/j.1469-7998.1993.tb05349.x

Watwood, S. L., Miller, P. J. O., Johnson, M., Madsen, P. T., & Tyack, P. L. (2006). Deep-diving foraging behaviour of sperm whales (Physeter macrocephalus). *Journal of Animal Ecology*, *75*(3), 814–825. https://doi.org/10.1111/j.1365-2656.2006.01101.x

Wood, C. M. (2003). *Environmental impact assessment: A comparative review* (Vol. 2nd). Harlow: Prentice Hall.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton, USA: Chapman and Hall/CRC.

Zar, J. H. (2010). *Biostatistical analysis* (5th ed.). London: Pearson Education Ltd.

Zuur, A. F., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed effects models and extensions in ecology with R*. New York, USA: Springer.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.