HAL open science

Modelling the spatial distribution of cetaceans in New Zealand waters

Fabrice Stephenson, Kimberly Goetz, Ben R. Sharp, Theophile L. Mouton, Fenna L. Beets, Jim Roberts, Alison B. Macdiarmid, Rochelle Constantine, Carolyn J. Lundquist

To cite this version:

Fabrice Stephenson, Kimberly Goetz, Ben R. Sharp, Theophile L. Mouton, Fenna L. Beets, et al.. Modelling the spatial distribution of cetaceans in New Zealand waters. Diversity and Distributions, 2020, 26 (4), pp.495–516. 10.1111/ddi.13035 . hal-03411065

HAL Id: hal-03411065

https://hal.umontpellier.fr/hal-03411065v1

Submitted on 1 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Check for updates

Received: 16 June 2019 | Revised: 15 December 2019 | Accepted: 20 December 2019

DOI: 10.1111/ddi.13035

BIODIVERSITY RESEARCH

Diversity and Distributions WILEY

Modelling the spatial distribution of cetaceans in New Zealand waters

Fabrice Stephenson[1] | Kimberly Goetz[2] | Ben R. Sharp[3] | Théophile L. Mouton[4] | Fenna L. Beets[1] | Jim Roberts[2] | Alison B. MacDiarmid[2] | Rochelle Constantine[5,6] | Carolyn J. Lundquist[1,6]

[1]National Institute of Water and Atmospheric Research (NIWA), Hamilton, New Zealand

[2]National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand

[3]Fisheries New Zealand, Ministry for Primary Industries, Wellington, New Zealand

[4]Marine Biodiversity, Exploitation, and Conservation (MARBEC), UMR 9190, Université de Montpellier, Montpellier, France

[5]School of Biological Sciences, University of Auckland, Auckland, New Zealand

[6]Institute of Marine Science, University of Auckland, Auckland, New Zealand

Correspondence
Fabrice Stephenson, National Institute of Water and Atmospheric Research (NIWA), Hamilton, New Zealand.
Email: Fabrice.Stephenson@niwa.co.nz

Funding information
NIWA Coasts and Oceans Programme; New Zealand Ministry of Primary Industries, Grant/Award Number: PR2O1401

Editor: Juliano Sarmento Cabral

Abstract

Aim: Cetaceans are inherently difficult to study due to their elusive, pelagic and often highly migratory nature. New Zealand waters are home to 50% of the world's cetacean species, but their spatial distributions are poorly known. Here, we model distributions of 30 cetacean taxa using an extensive at-sea sightings dataset (n > 14,000) and high-resolution (1 km^2) environmental data layers.

Location: New Zealand's Exclusive Economic Zone (EEZ).

Methods: Two models were used to predict probability of species occurrence based on available sightings records. For taxa with <50 sightings (n = 15), Relative Environmental Suitability (RES), and for taxa with ≥50 sightings (n = 15), Boosted Regression Tree (BRT) models were used. Independently collected presence/absence data were used for further model evaluation for a subset of taxa.

Results: RES models for rarely sighted species showed reasonable fits to available sightings and stranding data based on literature and expert knowledge on the species' autecology. BRT models showed high predictive power for commonly sighted species (AUC: 0.79–0.99). Important variables for predicting the occurrence of cetacean taxa were temperature residuals, bathymetry, distance to the 500 m isobath, mixed layer depth and water turbidity. Cetacean distribution patterns varied from highly localised, nearshore (e.g., Hector's dolphin), to more ubiquitous (e.g., common dolphin) to primarily offshore species (e.g., blue whale). Cetacean richness based on stacked species occurrence layers illustrated patterns of fewer inshore taxa with localised richness hotspots, and higher offshore richness especially in locales of the Macquarie Ridge, Bounty Trough and Chatham Rise.

Main conclusions: Predicted spatial distributions fill a major knowledge gap towards informing future assessments and conservation planning for cetaceans in New Zealand's extensive EEZ. While sightings datasets were not spatially comprehensive for any taxa, these two best available approaches allow for predictive modelling of both more common, and of rarely sighted, cetacean species with limited available information.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2020 The Authors. *Diversity and Distributions* published by John Wiley & Sons Ltd.

KEYWORDS
boosted regression tree models, cetacean distribution, New Zealand, relative environmental suitability models, spatial management, species distribution models

1 | INTRODUCTION

Cetaceans are distributed throughout the world's oceans, coasts and some river systems predominantly feeding on zooplankton, fishes and squids. Some species are inherently difficult to study due to their brief periods of time at the sea surface, offshore habitat use, and/or elusive behaviour. As a result, our understanding of cetaceans is over-represented by more accessible (e.g., coastal), species such as bottlenose dolphins (*Tursiops truncatuus*) and humpback whales (*Megaptera novaeangliae*). The distribution, range and behaviour of most species is poorly known and ~40% are considered "data deficient" by the IUCN Red List (Davidson et al., 2012; IUCN, 2019). The impacts of natural and anthropogenic stressors may be highest in areas which harbour large numbers of species. Pompa, Ehrlich, and Ceballos (2011) identified nine areas around the world that contained the highest richness of marine mammal species, all of which were associated with oceanographic upwelling where cold and warm water mixing favours high productivity. Likewise, a study by Kaschner, Tittensor, Ready, Gerrodette, and Worm (2011) predicted marine mammal richness to be highest in temperate waters of both hemispheres with distinct hotspots around New Zealand, Japan, Baja California, the Galapagos Islands, the Southeast Pacific, and the Southern Ocean.

Numerous studies have documented the impact of increased anthropogenic activities such as underwater noise, pollution, and overexploitation of prey species on cetacean species (e.g., Read, 2008; Tyack et al., 2011; Jepson et al., 2016; Porter and Lai, 2017; Pirotta et al., 2018). Using predictive spatially explicit models, Davidson et al. (2012) produced global risk maps to determine which marine mammal species are likely to be most vulnerable to increasing anthropogenic activities. While most at-risk marine mammals were in agreement with those listed by the IUCN, an additional 15 species were identified in the Indo-Pacific, around South Africa, New Zealand, Argentina, and along western coasts of South America and Central Africa, suggesting that 37% of marine mammal species are at risk of extinction (Davidson et al., 2012).

New Zealand's Exclusive Economic Zone (EEZ) which extends over 4 million km^2 is a recognised global cetacean diversity hotspot (Davidson et al., 2012; Kaschner et al., 2011; Pyenson, 2011). Mainland New Zealand has a long and narrow landmass extending across a wide latitudinal range (≈35–48°S), resulting in a diversity of environmental conditions in its surrounding waters (Bradford-Grieve, Lewis, & Stanton, 1991; Leathwick, Elith, Francis, Hastie, & Taylor, 2006; Stevens, O'Callaghan, Chiswell, & Hadfield, 2019). The two main islands are surrounded by a vast continental shelf, populated by shallower shelf areas and deep ocean trenches. The dominant oceanographic feature is the Subtropical Front: a highly productive zone of mixing between higher salinity, nutrient poor, warm, northern waters, and lower salinity, nutrient rich, cold, southern water, which creates ideal environments for preferred cetacean prey species (many plankton, cephalopod and fish species; Bradford-Grieve et al., 2006; Leathwick et al., 2006).

There are 47 cetacean species, subspecies and/or ecotypes including resident, migrant or vagrant taxa in New Zealand waters—seven of these are listed as "Critically Endangered" (Māui dolphin, *Cephalorhynchus hectori maui*) or "Endangered" (Hector's dolphin (*Cephalorhynchus hectori hectori*), humpback whale (*Megaptera novaeangliae*), fin whale (*Balaenoptera physalus*), sei whale (*Balaenoptera borealis*), pygmy blue whale (*Balaenoptera musculus brevicauda*), Antarctic blue whale (*Balaenoptera musculus intermedia*)), under the IUCN threat classification system (IUCN, 2001) and 28 are considered 'Data Deficient' and unable to be classified (IUCN, 2019). All cetaceans in New Zealand's territorial seas and EEZ are protected under national law by the New Zealand Marine Mammals Protection Act (1978), which mandates that all physical or habitat disturbances to the animals must be avoided or mitigated. However, little is known about the distribution and habitat use patterns of cetaceans in the seas surrounding New Zealand, especially for those species that inhabit offshore waters. This general paucity of spatial information is a major limitation with respect to the management of potential threats to cetaceans, e.g., fishing, mineral extraction industries and other threats that are heterogenous in space (Baker et al., 2019).

Despite the fact that 50% of the world's cetacean species are known to occur in New Zealand waters, large-scale regular ship- or aerial-based cetacean distribution and abundance surveys of the EEZ are logistically and financially prohibitive. Prior studies have provided some distributional predictions based on limited available data, for example Thompson et al. (2013) relied solely on strandings data to predict beaked whales distributions, and historical whaling records were used to estimate southern right whale distributions (Jackson et al., 2016; Torres et al., 2013). Some regional-scale distributions have been developed for species or locations with high abundance or of particular conservation interest, such as the Hauraki Gulf (Dwyer, Clement, Pawley, & Stockin, 2016), Kermadec Islands (Duffy, Baker, & Constantine, 2015), West Coast of the South Island (Bräger & Schneider, 1998) and/or for Hector's and Māui dolphins (Hamner, Pichler, Heimeier, Constantine, & Baker, 2012). More recently, species sighting records have been combined with environmental predictor variables to provide estimates of species distributions using both complex modelling approaches, e.g., BRTs used to predict Māui dolphin distribution (Derville, Constantine, Baker, Oremus, & Torres, 2016), and southern right whale distribution throughout New Zealand (Torres et al., 2013) and in the Auckland Islands (Rayment, Dawson, & Webster, 2015). Simpler approaches, for example, RES as described by Kaschner, Watson, Trites, and Pauly (2006), have

been used for the predictions of global home range estimates of the majority of cetacean species.

Knowledge gaps about the spatial and temporal distribution of many cetacean species in New Zealand waters can be filled using species distribution models to account for data uncertainties while still providing results to inform management decisions (Anderson et al., 2016). Species distribution models have become a reliable and recognised method of predicting species' probability of occurrence and are an integral part of resource management and conservation biology (Elith et al., 2006; Guisan & Thuiller, 2005). Spatial information, such as from opportunistically collected cetacean sightings (Derville, Torres, Iovan, & Garrigue, 2018), can be used to model a species' ecological niche based on the assumption that the distribution of known encounters reflects the species' environmental preferences (Guisan & Zimmermann, 2000; Hirzel, Lay, Helfer, Randin, & Guisan, 2006). Our study combines functionally relevant, high-resolution environmental data (1 km^2 grid resolution) across New Zealand's EEZ and a large database of opportunistically collected at-sea cetacean sighting records (n = 14,513), to predict probability of occurrences for 30 cetacean taxa and for a subset of these taxa spatially explicit estimates of uncertainty; this information is crucial for conservation and marine spatial planning.

2 | METHODS

2.1 | Study area

The study area extends over 4.2 million km^2 of the South Pacific Ocean within the New Zealand EEZ (≈25–57°S; 162°E–172°W; Figure 1a). Feature names used in the text are shown in Figure 1a.

2.2 | Biological data

At-sea cetacean sightings records of 30 cetacean species, subspecies and species complexes, collected over the period 1970–2017 were collated from multiple databases (Table A1 in Appendix S1 and further information in Appendix S1). A database requested from the Department of Conservation (DOC) contained sightings originating from a variety of sources (the general public, seismic vessels, boat charters, scientific surveys, fishing vessels, aircraft, and New Zealand ferries) while the other databases were privately maintained. In New Zealand, most sightings are reported to DOC either using a sighting form available online (https://www.doc.govt.nz/marine-mammal-sighting-form) or a spreadsheet containing the same information. Recently, DOC has put controls in place to identify basic errors in sighting reports but they have not been applied to older data. Therefore, the combination of databases being maintained by other organizations or scientists and the lag time between DOC receiving sightings data and the incorporation into their official database, it was important to compare databases to ensure that the most comprehensive sightings database possible was used in this analysis. We included data from privately maintained databases only if they were not already incorporated into the more comprehensive DOC database. The data were further quality controlled to remove any errors prior to analyses. This included removing ~6,000 records that: lacked species identification or location, were located on land, were located outside the New Zealand EEZ and were duplicated within and between databases. Following quality control, a total of 14,513 cetacean sighting records across 30 cetacean taxa were retained for final analysis, however these were unequally distributed between cetacean taxa (Table 1) and across the study region (Figure 1a). Details on at-sea cetacean sightings can be found in Appendix S1.

Depending on the number of records available for each taxa, different analyses were undertaken to estimate distributions. For those taxa with fewer than 50 recorded sightings, Relative Environmental Suitability (RES) models (Kaschner et al., 2006) were used to predict probability of occurrence (white shading in Table 1). For taxa with more than 50 recorded sightings, Boosted Regression Tree (BRT) models (Elith et al., 2006) were fitted to predict probability of occurrence (grey shading in Table 1). To predict probability of occurrence, BRT models require locations of both presences (sightings records) and absences. Here, true absences (i.e., location records for where cetacean species were not sighted) were not available. Although BRTs can be fitted with "background data" (also referred to as pseudo absences), here we opted to fit individual species models with presences of the other cetacean species within the database that were not being modelled. That is, if only a single species was sighted at a particular location, other species were logically absent. Here we refer to this as "relative absence" although it has also been referred to as "target-group background data" (Phillips et al., 2009). Although not as robust as true absences, the use of relative absences have been shown to improve average performance for regression based models compared to using background data, especially when the relative absences are part of the same broad biological group and have been collected using similar methods with the same sampling biases (as was the case here; Anderson, 2003; Phillips et al., 2009). To further minimise the effect of spatial sample-selection bias in the presence-relative absence data, all cetacean records were weighted by ($1/n$) where n was the number of presence records (of the species being modelled) in each 1 km^2 grid cell (Anderson et al., 2016).

We grouped sightings of the long-finned (*Globicephala melas*) and short-finned pilot whales (*G. macrorhynchus*) into the species complex "pilot whales", but recognise that most sightings where species identity was confirmed are of long-finned pilot whales (Oremus et al., 2009). Genetics and expert sightings data to date are dominated by pygmy blue whale records (*Balaenoptera musculus brevicauda*) but Antarctic blue whales (*B. m. musculus*) are also present in New Zealand waters (Barlow et al., 2018; Miller et al., 2014); we collectively refer to them as blue whales. We have chosen to analyse the Hector's dolphin (*Cephalorhynchus hectori hectori*) and Māui dolphin

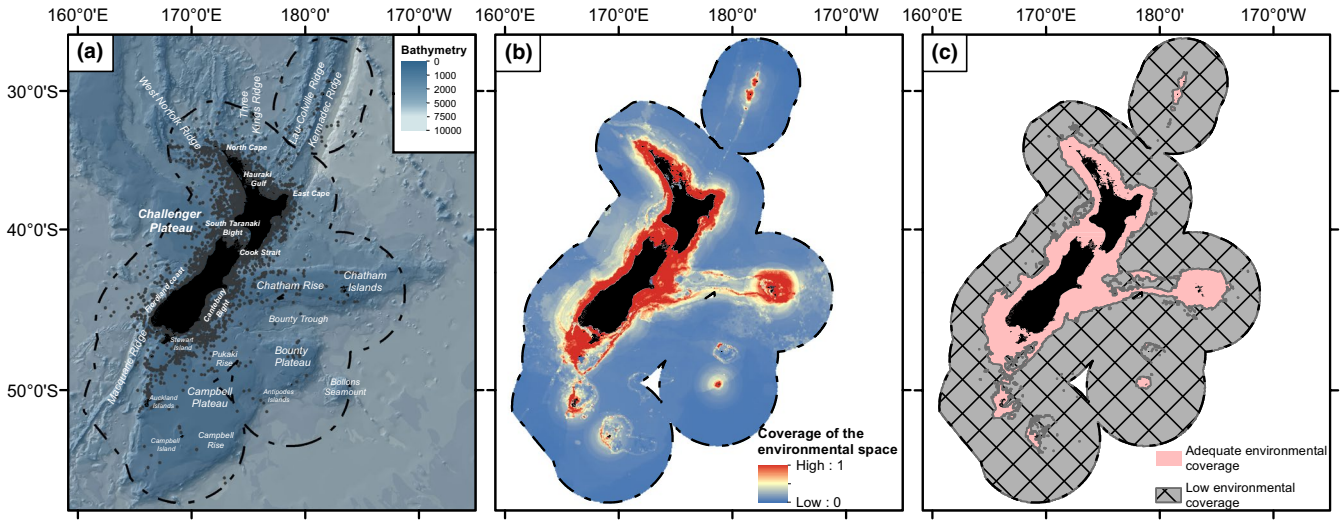

FIGURE 1 Map of the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line), bathymetry and species' sightings (black dots) and feature names used throughout the text (white) (a). Predicted environmental coverage depicting the confidence that can be placed in the predictions, ranging from low (i.e., no samples in the dataset with those environmental conditions) to high (i.e., many samples with those environmental conditions) within the New Zealand EEZ (b) and and subjectively defined areas of low environmental coverage (criss-cross black lines) (c)

(*C. h. maui*) data separately as there is a reasonably clear distinction in geographic range between these endemic sub-species (Hamner et al., 2012).

2.3 | Environmental data

New Zealand's large EEZ encompasses a diverse range of environmental conditions (Bradford-Grieve et al., 2006; Stevens et al., 2019). To capture this variability, 14 high resolution gridded environmental predictor variables (1 km^2), were collated and imported into Esri ArcGIS (version 10.6; described in Table 2; figures are shown in the Appendix S2). Variables were selected based on prior information with respect to their relationships with cetacean occurrence and distribution. While it is unlikely that the majority of these variables directly affect cetacean occurrence and distribution, physical processes and oceanographic features such as bathymetry, thermal layers, shelf breaks, and productivity are known to aggregate prey, in turn attracting cetaceans for foraging (Bluhm, Coyle, Konar, & Highsmith, 2007; Etnoyer et al., 2006; Tynan et al., 2005). In addition, certain environmental features may provide ideal habitat for cetaceans at different life history stages, e.g., cold, calm harbour waters at the Auckland Islands for southern right whale calving (Patenaude, Baker, & Gales, 1998). Although most of the chosen ocean climate variables were static (e.g., bathymetry), several variables were dynamic in time, representing mean monthly statistics (e.g., mixed layer depth, "temporal resolution" column in Table 2).

Although BRT modelling is reasonably robust to correlated variables (Elith & Leathwick, 2009; Guisan et al., 2013), the use of highly correlated variables generally provides only minimal improvement in predictive accuracy, and complicates interpretation of model outcomes (Leathwick et al., 2006). Variables used in the analysis (Table 2) had acceptable levels of correlation (Pearson's correlation coefficient ≤ 0.75, Fig A2-1 in Appendix S2 as per methodology in Dormann et al., 2013).

Prior to BRT model fitting, values for each environmental variable were extracted for cetacean sighting location by overlaying these records onto each of the environmental variable layers using the "raster" package in R (Hijmans & van Etten, 2012). For mean monthly environmental variables, recorded dates of cetacean sightings were used to extract respective values from the month the observation was made.

2.4 | Relative environmental suitability models

RES models were used to predict the geographical ranges of species <50 sightings records using basic descriptive data that are available for most species, including those for which few (or no) recorded locations are available (Kaschner et al., 2006). Here, cetacean sightings data were not used as inputs in the model but were used as a visual validation only.

Three environmental variables were selected to describe species' geographic ranges in RES models following methods described in Kaschner et al. (2006): sea surface temperature (SST), water depth (Bathy) and distance to shore (DistShore). The RES method does not require cetacean location records, but rather relies on the generic relationships between species and each of the environmental layers from the scientific literature and expert opinion. The relationships between species and the three environmental variables are described using a trapezoidal response curve based on four parameters: Min_A, Min_P, Max_P and Max_A. MinA and Max_A refer to absolute

TABLE 1 Number of cetacean sighting records per species, subspecies or species complexes included in the analyses

Species/subspecies/species complex names	Species/subspecies	Number of sightings records
Blainville's beaked whale	*Mesoplodon densirostris*	1
Dwarf minke whale	*Balaenoptera acutorostrata*	1
Spectacled porpoise	*Phocoena dioptrica*	1
Striped dolphin	*Stenella coeruleoalba*	1
Andrew's beaked whale	*Mesoplodon bowdoini*	2
Hourglass dolphin	*Lagenorhynchus cruciger*	2
Pygmy sperm whale	*Kogia breviceps*	2
Southern bottlenose whale	*Hyperoodon planifrons*	4
Risso's dolphin	*Grampus griseus*	5
Shepherd's beaked whale	*Tasmacetus shepherdi*	5
Cuvier's beaked whale	*Ziphius cavirostris*	7
Gray's beaked whale	*Mesoplodon grayi*	9
Southern right whale dolphin	*Lissodelphis peronii*	27
False killer whale	*Pseudorca crassidens*	28
Arnoux's beaked whale	*Berardius arnuxii*	31
Minke whale	*Balaenoptera acutorostrata*	57
Fin whale	*Balaenoptera physalus*	61
Sei whale	*Balaenoptera borealis*	70
Blue whale (spp. and sub spp.)	*Balaenoptera musculus musculus* *Balaenoptera m. brevicauda*	354
Southern right whale	*Eubalaena australis*	477
Sperm whale	*Physeter macrocephalus*	497
Bottlenose dolphin	*Tursiops truncatus*	498
Killer whale	*Orcinus orca*	569
Bryde's whale	*Balaenoptera edeni brydei*	593
Humpback whale	*Megaptera novaeangliae*	629
Pilot whale (2 spp.)	*Globicephala melas* *Globicephala macrorhynchus*	679
Dusky dolphin	*Lagenorhynchus obscurus*	823
Māui dolphin	*Cephalorhynchus hectori maui*	1,051
Hector's dolphin	*Cephalorhynchus hectori hectori*	3,688
Common dolphin	*Delphinus delphis*	4,411

Note: White shading indicates taxa for which Relative Environmental Suitability models were run; Grey shading indicates taxa for which Boosted Regression Tree probability of occurrence models were fitted.

minimum and maximum variable ranges (i.e., beyond these values, the species are not expected to occur), while Min_P and Max_P describe the "preferred" range, in terms of habitat usage of a given species (Kaschner et al., 2006).

Min_A, Min_P, Max_P and Max_A were initially defined for each species using values from global marine mammal species distribution models presented in Kaschner et al. (2006) (and references therein) and updated to a New Zealand context by a literature search and expert advice (six national and international experts on cetacean biology and ecology) for New Zealand specific species–habitat relationships (all the values used and a qualitative description of the ranges for each species are provided in Appendix S3). For each species, RES scores for the three environmental variables (SST, Bathy, DistShore) were produced by transforming the gridded data layers of the variables following the trapezoidal response curve. For each species, final RES scores were produced by multiplying the three suitability layers assigned to the individual attributes (e.g., SST, Bathy, DistShore) resulting in spatial estimates of environmental suitability ranging from 0 (not present) to 1 (highly representative of the species preferred or overall habitat range). It was not possible to quantitatively evaluate RES model performance because there were too few at-sea sightings for the cetacean taxa distributions modelled using this method. To visually evaluate mapped RES predictions a further dataset of cetacean beach strandings (obtained from the New Zealand Department

TABLE 2 Environmental variables used as predictors in Boosted Regression Tree analyses

Variable abbreviation	Variable name	Temporal resolution	Unit	Description	Source
Bathy	Bathymetry	Static	m	Depth at the seafloor was interpolated from contours generated from various sources, including multi-beam and single-beam echo sounders, satellite gravimetric inversion, and others (Mitchell et al., 2012).	Mitchell et al. (2012)
BedDist	Benthic sediment disturbance	Static	unitless	Combination of seabed orbital velocities (estimates the average mixing at the seafloor as a consequence of orbital wave action, calculated from a wave climatology derived hindcast [1979–1998] of swell-wave conditions in the New Zealand [NZ] region[Gorman, Bryan, & Laing, 2003]) and friction velocity for seabed types (based on grain size). Benthic sediment disturbance from wave action was assumed to be zero where depth ≥ 200 m.	Leathwick et al. (2012)
Chl*A*	Chlorophyll-a concentration	Monthly mean	mg m^{-3}	A proxy for the amount of photosynthetic plankton, or phytoplankton, present in the ocean. Estimated using quasi-analytic inversion algorithm applied to MODIS-Aqua data. Results were calculated based on long-term (2002–2017) average values of phytoplankton absorption aph(555) at 500 m spatial resolution.	NIWA unpublished; Based on processing described in Pinkerton, Gall, Wood, and Zeldis (2018)
Dist.Iso500	Distance to 500 m isobath	Static	km	The 500 m bathymetric contour was used to denote the shelf break. Distance from this isobath was calculated using the spatial analysis extension in ArcGIS.	NIWA, unpublished
Dist.Shore	Distance to shore	Static	km	Using a NIWA sourced polygon of the New Zealand coastline, distance from shore was calculated using the spatial analysis extension in ArcGIS.	NIWA, unpublished
DOM	Coloured dissolved organic matter (CDOM)	Static	Indicative of CDOM absorption at 440 nm ag(440) (m^{-1})	Detrital absorption at 440 nm, including due to coloured dissolved organic matter (CDOM) and particulate detrital absorption. Estimated using quasi-analytic inversion algorithm applied to MODIS-Aqua data. Results were calculated based on long-term (2002–2017) average values of detrital absorption coefficient ag(443) at 500 m spatial resolution	NIWA unpublished; Based on processing described in Pinkerton et al. (2018)
Kpar	Diffuse downwelling attenuation	Monthly mean	m^{-1}	Attenuation of broadband irradiance (Photosynthetically Available Radiation, PAR) with depth. Estimated using quasi-analytic inversion algorithm applied to MODIS-Aqua data. Results were calculated based on long-term (2002–2017) average values at 500 m spatial resolution.	NIWA unpublished; Based on processing described in Pinkerton et al. (2018)
MLD	Mixed layer depth	Monthly mean	m	The depth that separates the homogenized mixed water above from the denser stratified water below.	Calculated from the CARS climatology, NIWA, unpublished
Slope	Slope	Static	Degree	Bathymetric slope was calculated from bathymetric depth and is the degree change from one depth value to the next.	NIWA, unpublished
SST	Sea surface temperature	Monthly mean	°C	MODIS-Aqua SST product, calculated as long-term (2002–2017) average values at 1,000 m resolution.	NIWA, unpublished; Based on processing described in Pinkerton et al. (2018)
TC	Tidal Current speed	Static	ms^{-1}	Maximum depth-averaged (NZ bathymetry) flows from tidal currents calculated from a tidal model for New Zealand waters (Walters, Goring, & Bell, 2001)	Leathwick et al. (2012)

(Continues)

TABLE 2 (Continued)

Variable abbreviation	Variable name	Temporal resolution	Unit	Description	Source
TempRes	Temperature residuals	Static	°C	Residuals from a GLM relating bottom water temperature to depth using natural splines—this highlights areas where average temperature is higher or lower than would be expected for any given depth. Positive values indicate waters of subtropical origin and to the west and north of New Zealand. Negative values indicate cool waters of subantarctic origin and are widespread east of the southern South Island and on the southern flanks of the Chatham Rise (see Figure A 2-13).	Leathwick et al. (2006)
Turb	Turbidity	Monthly mean	NTU	Optical backscatter as measured by turbidity sensor. Estimated using quasi-analytic inversion algorithm applied to MODIS-Aqua data. Result calculated based on long-term (2002–2017) average values of particulate backscatter bbp(555) at 500 m resolution, converted to normalised turbidity units (NTU) using in situ turbidity measurements in the New Zealand coastal zone.	NIWA unpublished; Based on processing described in Pinkerton et al. (2018)
VGPM	Productivity Model	Monthly mean	$mgCm^{-2}d^{-1}$	Provides estimates of surface water primary productivity based on the Vertically generalized productivity model of Behrenfeld and Falkowski (1997). Net primary productivity by phytoplankton (mean daily rate of water column carbon fixation) is estimated as a function of merged remotely sensed chlorophyll concentration, irradiance, and photosynthetic efficiency estimated from remotely sensed Sea-Viewing Wide-Field-of-view Sensor (SeaWIFS) and MODIS-Aqua satellite imagery (M. Pinkerton, NIWA, personal communication)	NIWA unpublished; Oregon State University (www.science.oregonstate.edu/ocean.productivity/)

Note: Variable abbreviations, full names, temporal resolution, units and description are provided. Figures of the environmental variables used are shown in the Appendix 2.

of Conservation which included 3,410 "freshly dead" or "alive" stranding's records of 44 cetacean species, collected over the period 1900–2018) were used along with the recorded at-sea sightings (<50 for these taxa).

2.5 | Boosted Regression Tree model fitting and evaluation

For those species with sufficient number of records (>50 recorded sightings), relationships between cetacean presence/relative absence and environmental variables were investigated using BRT models. BRT modelling combines many individual regression trees (models that relate a response to their predictors by recursive binary splits) and boosting (an adaptive method for combining many simple models to give improved predictive performance) to form a single ensemble model (Elith, Leathwick, & Hastie, 2008). Detailed descriptions of the BRT method are available in Ridgeway (2007) and Elith et al. (2008). All statistical analyses were undertaken in R (R Core Team, 2013) using the "Dismo" package (Hijmans, Phillips, Leathwick, & Elith, 2017).

BRT models were fitted with a Bernoulli error distribution, a tree complexity of 5, a learning rate between 0.01–0.0001 (with the parameter selected so as to fit between 1,000 and 3,000 trees for each species' model), a bag fraction of 0.6 and random 10-fold cross evaluation following recommendations from Elith et al. (2008) and Leathwick et al. (2006).

For all individual cetacean taxa with >50 recorded sightings, BRT models were initially fitted using all available environmental variables. The cross-validation process ensures that models are parsimonious, however, over-fitting can also occur by including more predictor variables than necessary (Leathwick et al., 2006). To reduce the risk of overfitting, the global models (those with all predictor variables included) for each species in turn were subjected to a simplification process whereby environmental variables were removed from the models, one at a time, using the "simplify" function (Elith et al., 2006). This simplification process firstly assesses the relative contributions of each variable in terms of deviance explained, with the lowest contributing variables removed from the model, before the model is refitted with the remaining environmental variables. The change in deviance explained that resulted from removing the variable was then examined and the process repeated until all variables were sequentially removed. The final models were created by refitting the model with a reduced variable set that balanced the deviance explained with a reasonable number of predictor variables (Table 5).

BRT models were assessed using cross-validated measures of model performance (Compton, Morrison, Leathwick, & Carbines,

TABLE 3 Number of positive occurrences within the NOMAD database aggregated to a 1 km^2 for those species with ≥50 positive species records (presences)

Common name	Number of positive occurrences (aggregated to 1 km^2)
Bottlenose dolphin	146
Killer whale	58
Dusky dolphin	320
Hector's dolphin	1,648
Common dolphin	1,776

Note: When aggregated to 1 km^2, there were 15,900 cells where no species were recorded (absences) while observers were on-effort.

2012; Elith et al., 2008). Model performance measures included the deviance explained and the area under the receiver operating characteristic curve (AUC). The explained deviance provides a measure of the goodness-of-fit between the predicted and raw values (total deviance; Compton et al., 2012). AUC measures the model's ability to discriminate between presence and absence points. AUC ranges from 0 to 1, where a value of 1 indicates that the presences and absences are perfectly discriminated, while a value of ≤0.5 indicates that discrimination is no better than random chance (Elith et al., 2006). The relative influence of each environmental variable in the models was the number of times it was selected for splitting, weighted by the squared improvement to the model as a result of each split (using in-bag data; Friedman & Meulman, 2003). The association between species occurrence and the four most influential environmental predictor variables was illustrated using partial dependence plots (i.e., predicted response curve of species probability of occurrence across the gradient of the variable of interest when all other variables are held at their means).

BRT models were bootstrapped 100 times for each of the 15 cetacean taxa with >50 sightings. A random 'training' sample of the presence-relative absence records was drawn with replacement. That is, 75% of the presence data and twice as many relative absences were randomly selected from the database, and a model was constructed with the same settings as the original. Relative absences were not selected in locations of presences (the minimum distance between presence and relative absence was 1 km). The bootstrapping process was repeated 100 times, and at each iteration, predictions were made to the "evaluation" data, i.e., the remaining 25% of the presence data and twice as many randomly selected relative absences from the remaining absences, allowing model fits to be examined both on the training and evaluation data. Fin, minke and sei whale models did not always converge when bootstrapped due to the low number of samples. Models were therefore only run once for these taxa.

2.6 | Assessing model uncertainty

For each of the 15 cetacean taxa with >50 sightings, bootstrapped BRT models were predicted geographically using the mean of the monthly mean and static environmental predictor variables to a 1 km^2 grid. For each taxa, mean probability of occurrence and a spatially explicit measure of uncertainty (measured as the standard

TABLE 4 Cross-validated estimates of model performance for the bootstrapped BRT models fitted with presence/relative absence sightings per species

Taxa	Deviance explained (training data)	Deviance explained (evaluation data)	AUC (training data)	AUC (evaluation data)	NOMAD evaluation (AUC)
Fin whale	0.20	0.19	0.90	0.81	NA
Minke whale	0.18	0.25	0.88	0.79	NA
Sei whale	0.24	0.24	0.88	0.81	NA
Killer whale	0.18 ± 0.02	0.16 ± 0.02	0.79 ± 0.01	0.79 ± 0.02	0.68
Bottlenose dolphin	0.20 ± 0.02	0.19 ± 0.02	0.81 ± 0.01	0.81 ± 0.02	0.71
Humpback whale	0.29 ± 0.02	0.28 ± 0.03	0.85 ± 0.01	0.85 ± 0.02	NA
Common dolphin	0.39 ± 0.00	0.39 ± 0.01	0.90 ± 0.01	0.90 ± 0.01	0.68
Pilot whale	0.43 ± 0.02	0.43 ± 0.03	0.91 ± 0.01	0.91 ± 0.01	NA
Sperm whale	0.46 ± 0.02	0.46 ± 0.04	0.92 ± 0.01	0.92 ± 0.01	NA
Bryde's whale	0.49 ± 0.02	0.50 ± 0.04	0.93 ± 0.01	0.93 ± 0.01	NA
Suthern right whale	0.53 ± 0.02	0.53 ± 0.04	0.94 ± 0.01	0.94 ± 0.01	NA
Dusky dolphin	0.55 ± 0.01	0.55 ± 0.03	0.95 ± 0.00	0.95 ± 0.01	0.91
Blue whale	0.58 ± 0.02	0.58 ± 0.05	0.95 ± 0.01	0.95 ± 0.01	NA
Māui dolphin	0.87 ± 0.01	0.88 ± 0.02	0.99 ± 0.00	0.99 ± 0.00	NA
Hector's dolphin	0.83 ± 0.01	0.83 ± 0.01	0.99 ± 0.00	0.99 ± 0.00	0.96

Note: Model performance was assessed using the proportion of the mean total deviance explained ± standard deviation (*SD*) and mean AUC (area under the receiver operating curve) (AUC) ± standard deviation for models fitted with the training data (75% of sightings records) and the evaluation data (25% sightings records). AUC values from independent evaluation using true presence/absence NOMAD records are provided for select species (NA: not available). Fin, minke and sei whales were not bootstrapped and therefore only the results from a single model are shown.

deviation of the mean [*SD*]) were calculated for each grid cell using the 100 bootstrapped BRT layers.

As an added measure of model uncertainty 'coverage of the environmental space by cetacean sightings records' was estimated (Smith, Duffy, Anthony, & Leathwick, 2013). The "environmental space" is the multidimensional space when each variable is treated as a dimension. Cetacean sightings records can be projected into this space, where some parts of this environmental space will contain many sighting records (and are therefore well covered by the sightings data) and other parts of this environmental space will contain few sighting records (and therefore the relationship between the environment and the sightings records are poorly understood resulting in potentially less certain predictions). For the latter where predictions are considered less reliable, these should be treated with more scepticism than areas that are well covered by sightings records (Smith et al., 2013). Here we followed methods described in Smith et al. (2013) to model the coverage of the environmental space.

The degree to which the environmental conditions of each predictive site was covered by the sightings records was quantified by randomly sampling 50,000 values from the environmental space (where no cetacean sightings were recorded) and assigning a "records" value of 0 to these, indicating that these were "absent" sample sites. These were combined with the true sightings records (n = 14,513), to which a "records" value of 1 was assigned ("present"). A BRT was then used to model the relationship between "absent" (random) records and "present" (true) records for the 14 environmental predictor variables, using a Bernoulli error distribution. Predictions using this model yielded estimates of the probability of a site occurring in each part of the environmental space. A learning rate that yielded 2,000 trees with an interaction depth of 2 was used (so that only pair-wise combinations of the environmental variables were considered). Predictions were then made spatially, generating values between 0 and 1 (where 0 indicated little understanding of the environmental space and 1 a perfect understanding), according to how well each cell was represented by the sighting records.

To further investigate the robustness of RES models, RES model predictions using sea surface temperature (SST), water depth (Bathy) and distance to shore (Dist.Shore) as variables were created for those species with sufficient number of records (>50 recorded sightings) and were visually compared with their respective BRT predictions. For those species for which sea surface temperature (SST), water depth (Bathy) and distance to shore (Dist.Shore) were important environmental variables (or covaried with other important variables, e.g., turbidity) RES predictions were deemed robust. For those species for which sea surface temperature (SST), water depth (Bathy) and distance to shore (Dist.Shore) were not important variables in the BRT models, the RES predictions were less representative. As with other correlative modelling, RES predictions for individual species will be reliant on the selection of relevant environmental variables. However, given the lack of information for rarely sighted cetacean species (<50 recorded sighting), sea surface temperature (SST), water depth (Bathy) and distance to shore (Dist.Shore) were thought to most likely capture their ecological niches.

2.7 | Independent evaluation of Boosted Regression Tree models

Independently collected presence and absence data were collected by the New Zealand Ministry for Primary Industries (MPI) inshore fisheries observers using "Trimble Nomad" GPS-based data loggers since 2009. Trained fishery observers were required to conduct an inspection of the waters surrounding vessel approximately every 20 min. These data were collated by DOC into a single database (NOMAD, unpublished database, held by MPI/DOC), recording all on-effort periods and positive sightings of all cetacean species observed (i.e., on-effort events with no sightings can be considered true zeros for all species not observed). In this study, NOMAD records were used to validate the predicted probability of occurrence for models for which there were ≥50 positive species records (aggregated to 1 km^2) in the NOMAD database (Table 3; bottlenose dolphin, dusky dolphin, Hector's dolphin, common dolphin and killer whale). NOMAD presence/absence data were used to calculate the area under the receiver operating characteristic curve (AUC) by comparing the predicted values from the probability of occurrence models at locations of observed presence/absence.

2.8 | Estimating cetacean richness

As an example of the possible uses of the layers produced here, cetacean richness was estimated by stacking species distribution model predictions (S-SDM). That is, cetacean richness was estimated as the sum of the occurrence probability predictions (ranging from 0 to 1) from individual models (15 BRT models and 15 RES models; Calabrese, Certain, Kraan, & Dormann, 2014; Ferrier & Guisan, 2006). To reflect the lower prediction certainty associated with the coarser RES predictions, these were subjectively down-weighted by multiplying RES probability of occurrences by 0.25. For future applications the subjective weighting of RES layers could be further explored. The estimated distribution of cetacean richness therefore ranged from 0 to a theoretical maximum of 19 which was clipped to areas with adequate environmental coverage.

3 | RESULTS

3.1 | Boosted Regression Tree model performances

Based on model fit measures using cross-validated data (evaluation data), all cetacean occurrence models were considered useful (i.e., AUC > 0.75; Table 4). AUC scores (measuring the ability to discriminate between presence and relative absence points) ranged from 0.79 (killer whale) to 0.99 (Hector's and Māui dolphins) with a mean of 0.90 (Table 4). Models were able to explain between 16% (killer whale) and 88% (Māui dolphin) deviance in species presence/relative absence (Table 4). The model fit metrics between species' training data and evaluation data were in most cases very similar (Table 4) suggesting

TABLE 5 Variable contributions for bootstrapped BRT models fitted with presence/relative absence sightings per cetacean species

	Fin whale	Minke whale	Sei whale	Killer whale	Bottlenose dolphin	Humpback whale	Common dolphin	Pilot whale	Sperm whale	Bryde's whale	Southern right whale	Dusky dolphin	Blue whale	Māui dolphin	Hector's dolphin
Environmental Variables				Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean	Mean
Bathy	16.1	14.1	5.3	16.9 ± 0.4	12.1 ± 0.3		0.3 ± 0.0	25.3 ± 0.8	73.1 ± 0.3	3.0 ± 0.1	8.5 ± 0.6	6.9 ± 0.2	9.3 ± 0.3		4.3 ± 0.0
BedDist			3.6										4.2 ± 0.2		
ChlA	9.0	4.0	5.8	6.6 ± 0.2		7.9 ± 0.2	0.5 ± 0.0			6.7 ± 0.2		0.8 ± 0.0		1.0 ± 0.0	
Dist.Iso500	1.7	18.2	8.4	7.0 ± 0.2		8.5 ± 0.2	2.3 ± 0.0	2.2 ± 0.1	3.7 ± 0.1		2.6 ± 0.2	16.4 ± 0.2	11.1 ± 0.4	4.7 ± 0.2	3.3 ± 0.0
Dist.Shore	1.9	2.8	7.3	11.4 ± 0.3	11.9 ± 0.3	8.4 ± 0.2	17.1 ± 0.1	52.7 ± 0.8	2.8 ± 0.1		15.8 ± 0.8		48.8 ± 0.4		
DOM	17.5	33.5	5.0	3.3 ± 0.1			1.5 ± 0.0	6.5 ± 0.3	4.2 ± 0.1		6.8 ± 0.4	1.3 ± 0.1	1.9 ± 0.1		
Kpar	3.9					5.6 ± 0.2	1.5 ± 0.1	2.6 ± 0.1			7.6 ± 0.5	0.9 ± 0.0	4.0 ± 0.1	0.7 ± 0.0	
MLD	5.3	21.8	3.9	9.4 ± 0.2	9.6 ± 0.2	27.9 ± 0.4	0.7 ± 0.0		1.8 ± 0.1	1.6 ± 0.1	27.4 ± 0.8	3.4 ± 0.1	2.4 ± 0.1		29.5 ± 0.3
Slope	3.4		3.2	10.3 ± 0.2	10.7 ± 0.3	15.1 ± 0.3							3.2 ± 0.2		
SST	13.7		4.6	9.8 ± 0.3		11.1 ± 0.3	9.0 ± 0.1	5.3 ± 0.1	2.5 ± 0.1	6.2 ± 0.2	26.4 ± 1.0	10.5 ± 0.2	7.2 ± 0.2		10.9 ± 0.1
TC	2.6		6.1	6.6 ± 0.3			0.6 ± 0.0		5.5 ± 0.1	9.7 ± 0.3		1.4 ± 0.1		14.8 ± 0.2	
TempRes	5.1		6.2	18.6 ± 0.3	26.5 ± 0.3	15.5 ± 0.2	42.9 ± 0.1	2.0 ± 0.1	3.3 ± 0.1	51.7 ± 0.2	4.8 ± 0.3	55.3 ± 0.3	4.6 ± 0.2	72.7 ± 0.1	24.9 ± 0.1
Turb	12.0	5.6	40.7		15.3 ± 0.2		21.5 ± 0.1	3.5 ± 0.1		13.6 ± 0.3		1.8 ± 0.1	3.2 ± 0.2	6.1 ± 0.1	25.5 ± 0.3
VGPM	8.0				13.8 ± 0.2		2.2 ± 0.0		3.0 ± 0.1	7.6 ± 0.2		1.4 ± 0.1			1.6 ± 0.0

Note: The percent contribution (%) ± standard error (*SE*) is shown for each environmental variable (light grey cells represent low percent contribution; darker grey cells represent higher percent contributions). For those variables not used in the final BRT models, table cells are blank. Fin, minke and sei whales were not bootstrapped due to low sample size and therefore only the results from a single model are shown.

that models were not overly fitted to the training data. Standard deviations for each model fitting metric were low, providing evidence that models were consistently performing across bootstrap samples (Table 4). Model validation for a select number of species with independently collected presence/true absence data showed that all models had some predictive power (NOMAD evaluation AUC > 0.68), although in some cases this was substantially lower than for evaluation with non-independent data (e.g., common dolphin, AUC was 0.9 compared to 0.68 when using the NOMAD data; Table 4).

3.2 | Variable selection and contribution

The number of environmental predictor variables retained for species' models ranged from 6 to 13, with a median of 8 across all species (Table 5). The relative importance of each environmental variable (in terms of % contribution) varied across species (Table 5). The most consistently important variable for predicting the presence/relative absence of cetacean taxa was temperature residuals (TempRes) which was selected in 14 out of 15 models with an average contribution of 23.8% (Table 5). The next most important variables across all species' models, respectively, were bathymetry (Bathy), distance to the 500 m isobath (Dist.Iso500), mixed layer depth (MLD) and water turbidity (Turb)—each of which was selected in 13 of the 15 species' models (Table 5). Benthic sediment disturbance (BedDist) was seldom included in the models due to its poor predictive power relative to other environmental variables (Table 5). The locations of at-sea sightings for each species, the association between species occurrence and the four most influential environmental predictor variables is shown in partial dependence plots for each species in the Appendix S4. An example of interpretation of these plots for the common dolphin is also provided in the Appendix S4.

3.3 | Geographic predictions

Geographic predictions of the coverage of the environmental space by sightings records and a subset of predicted distribution of species' occurrences are presented here. Species' predicted occurrences were split into four categories with an example provided for each: species with restricted ranges predicted to occur in inshore waters (i.e., predominately < 25 km from the coastline), species predicted to occur in both inshore and offshore areas, species predicted to occur predominately in offshore waters for which more than 50 sightings records were available (BRT models), and species predicted to occur predominately in offshore waters for which few sightings records were available (RES models). We have presented a representative species for each of the geographic regions to show how the model functioned. Predicted distribution of species occurrence and associated spatially explicit uncertainty estimates for all other species are provided in Appendix S4. In addition, example R code for BRT, RES and "coverage of the environmental space" models are available in Appendix S5.

3.3.1 | Coverage of the environmental space

Geographic prediction of the coverage of the environmental space by the sighting records provided a spatially explicit indication of areas where, for all species' models, predictions were extrapolated beyond the environmental characteristics of the input data (Figure 1b). Areas where the environmental space was well covered by sighting records were predominately located within 200 km from shore around the North Island, South Island and the Chatham Islands, as well as smaller areas surrounding the Auckland, Campbell and Antipodes Islands and Kermadec Ridge (red areas in Figure 1b). Poorly covered areas included much of the study area further from shore (blue areas in Figure 1b). Here, we have recommended a subjectively defined threshold of 0.075 (i.e., in areas with lower values we have limited understanding of the environmental space and therefore predictions of species' distributions are less certain). Areas with values lower than this cut-off are shown in the criss-crossed grey areas in Figure 1c. In these poorly covered areas of the environmental space, geographic predictions from species' occurrence models should be treated with greater scepticism and caution (Elith & Leathwick, 2009).

3.3.2 | Inshore taxa

Species with restricted ranges predicted to occur primarily in inshore waters (≈25 km offshore) included: Hector's dolphin (presented below) and Māui dolphin (predicted species' distribution and associated uncertainty are available in Appendix S4).

Hector's dolphin

Hector's dolphins were predicted to be present very close to shore (no further than 50 km) along the coast of the South Island (Figure 2). The highest probability of occurrence was located off the east coast of South Island (Figure 2iv). Spatially explicit estimates of uncertainty (*SD*) were low across the study area (Figure A4-80 in Appendix S4). However, these uncertainty estimates increased slightly for those areas immediately outside Hector's dolphin core distribution (slightly further from the coast; Figure A4-80 in Appendix S4). Hector's dolphin records and model outputs fall within areas of the environmental space that are considered well sampled (Figure 2) and as such uncertainty layers are likely to provide an accurate representation of spatial uncertainty.

3.3.3 | Inshore–offshore taxa

Species predicted to occur in both inshore and offshore areas included: common dolphin (presented below) as well as Bryde's whale, fin whale, bottlenose dolphin, killer whale, humpback whale and dusky dolphin (predicted species' distribution and associated uncertainty are available in Appendix S4).

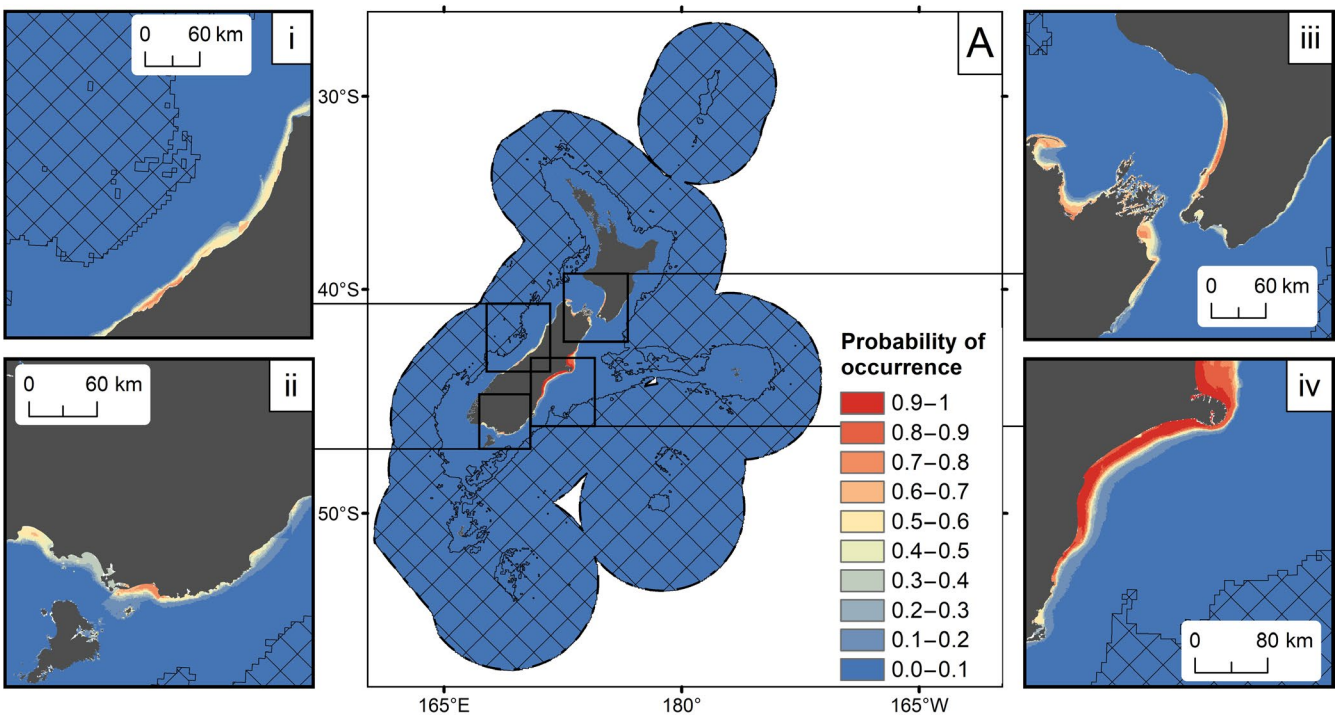

FIGURE 2 The predicted probability occurrence of Hector's dolphin (*Cephalorhynchus hectori hectori*) in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: (i) west coast of South Island including the Fiordland Coast; (ii) south of the South Island including Stewart Island/ Rakiura; (iii) south of the North Island and north of the South Island including Tasman and Golden Bays and Cook Strait; (iv) East of the South Island including Canterbury Bight. See Figure 1a, for place names

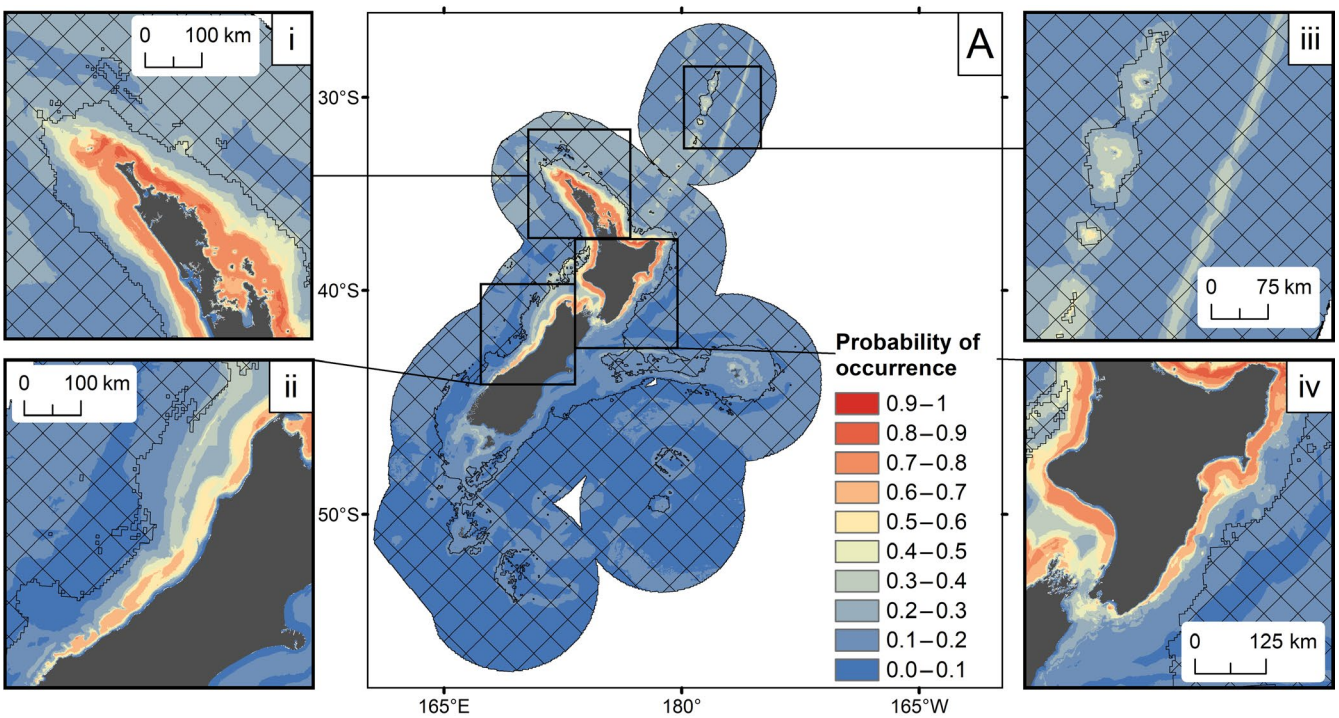

FIGURE 3 The predicted probability of common dolphin (*Delphinus delphis*) occurrence in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: (i) north of North Island including the North Cape and Hauraki Gulf; (ii) west coast of South Island including the Fiordland Coast; (iii) Kermadec islands, Lau-Colville Ridge and Kermadec Ridge; (iv) south of the North Island including the South Taranaki Bight and Cook Strait. See Figure 1a, for place names

Common dolphin

Common dolphins were largely predicted to occur within 50–100 km from shore (although not within the first 5 km) predominately in waters surrounding the North Island and the north and west coasts of the South Island (red areas in Figure 3). The highest probability of occurrence was predicted for areas between the North Cape and the Hauraki Gulf (Figure 3i) and from the Bay of Plenty to East Cape (Figure 3iv). Moderate predicted values of common dolphin occurrence (0.5–0.7) were observed in some areas further from shore (e.g., along the Kermadec Ridge, Figure 3iii). Spatially explicit estimates of uncertainty (measured here as *SD*) were low across the study area (Fig A4-84 in the Appendix S4). However, the interpretation of this uncertainty layer should be treated with some caution in areas with low environmental coverage values (further from shore) and where few occurrences were recorded (Figures 1 and 2).

3.3.4 | Offshore taxa—Boosted Regression Tree models

Species predicted to occur predominately in offshore waters for which there were sufficient records to fit BRT models included: blue whales (presented below) as well as minke whale, sei whale, southern right whale, sperm whale and pilot whales (predicted species distributions and associated uncertainty are available in Appendix S4).

Blue whales

Blue whales were predicted to occur across large offshore areas in the north of the study area (Figure 4). However, the highest probability of occurrence was predicted for areas in the South Taranaki Bight (Figure 4iv) in line with limited information of their distribution from prior studies. Spatially explicit estimates of uncertainty (measured here as *SD*) were low (*SD*: 0.0–0.1) in the southern part of the study area, but were moderate across the study area (*SD*: 0.1–0.2; Fig A4-42 in the Appendix S4). Moderate *SD* values (0.15–0.2) in the north of the study region coincided with areas with moderate—high probability of occurrence (Figure 4iii). In addition, as with most offshore species, aside for areas in the South Taranaki Bight, other areas with moderate—high probability of occurrence (North Cape and West Norfolk Ridge, northern section of the Challenger Plateau, Kermadec Islands, Lau-Colville Ridge and Kermadec Ridge, Figure 4) were in areas with low environmental coverage values for which predicted occurrences must be used with greater caution (Figure 2).

3.3.5 | Offshore taxa—Relative Environmental Suitability models

Species predicted to occur predominately in offshore waters for which little data were available (RES models) included: southern right whale dolphin (presented here) as well as Andrew's beaked whale, Arnoux's beaked whale, Blainville's beaked whale, Cuvier's beaked whale, dwarf minke whale, false killer whale, Gray's beaked whale, hourglass dolphin, pygmy sperm whale, Risso's dolphin, Shepherd's beaked whale, southern bottlenose whale, spectacled porpoise and striped dolphin (predicted species' distribution are available in Appendix S4).

RES models were used for rarely sighted species, the majority of which were predicted to have highest RES scores offshore. It was not possible to quantitatively evaluate RES model performance and therefore these predictions are considered to have lower certainty than for those species modelled using BRTs. However, visual comparison of recorded sightings and beach strandings data with the predicted RES scores suggests that for the majority of these rarely observed species, these distributions appeared reasonable (see RES predictions in Appendix S4). For example, predicted RES scores for southern right whale dolphin show that the highest RES scores (score of 0.9–1) covered extensive parts of the study area (Figure 5). Sightings records for southern right whale dolphin (n = 27) were located in a cluster offshore of the south-east coast of the South Island in areas with high RES values providing some evidence that this species has a preference for deeper, offshore waters (Figure 5). Further, the stranding records were located on both the North Island and South Island (as far north as the Bay of Islands), providing some evidence that this species may also use the offshore areas in the North Island despite no sightings recorded further north than the Taranaki Bight (Figure 5).

3.4 | Cetacean richness

Estimated cetacean richness in New Zealand ranged from 1 to 10 species km^{-2} (Figure 6). Broadly, offshore areas (>25 km from the coastline) had the highest predicted species richness, specifically, along the Macquarie Ridge (Figure 6ii), the western edge of the Bounty Trough and along the southern and northern edges of the Chatham Rise (Figure 6iv). Inshore (<25 km from the coastline) species richness was generally predicted to be lower (estimated cetacean richness between 1–4), although Fiordland coast, North Cape, South Taranaki Bight, Kaikoura Coast, Cook Strait, East Bay of Plenty and the Kermadec Islands all had regions of with moderate—high species richness (estimated cetacean richness: 5–8, Figure 6).

4 | DISCUSSION

Despite the high number of cetacean species observed, information on cetacean distributions in New Zealand waters is limited (Baker et al., 2019). Here, a comprehensive set of spatial information for 30 cetacean species, subspecies, and species complexes occurring in New Zealand waters was combined with high resolution environmental predictors to estimate species' occurrence across the New Zealand EEZ using two modelling methods. These predicted spatial distributions provide a step towards delivering

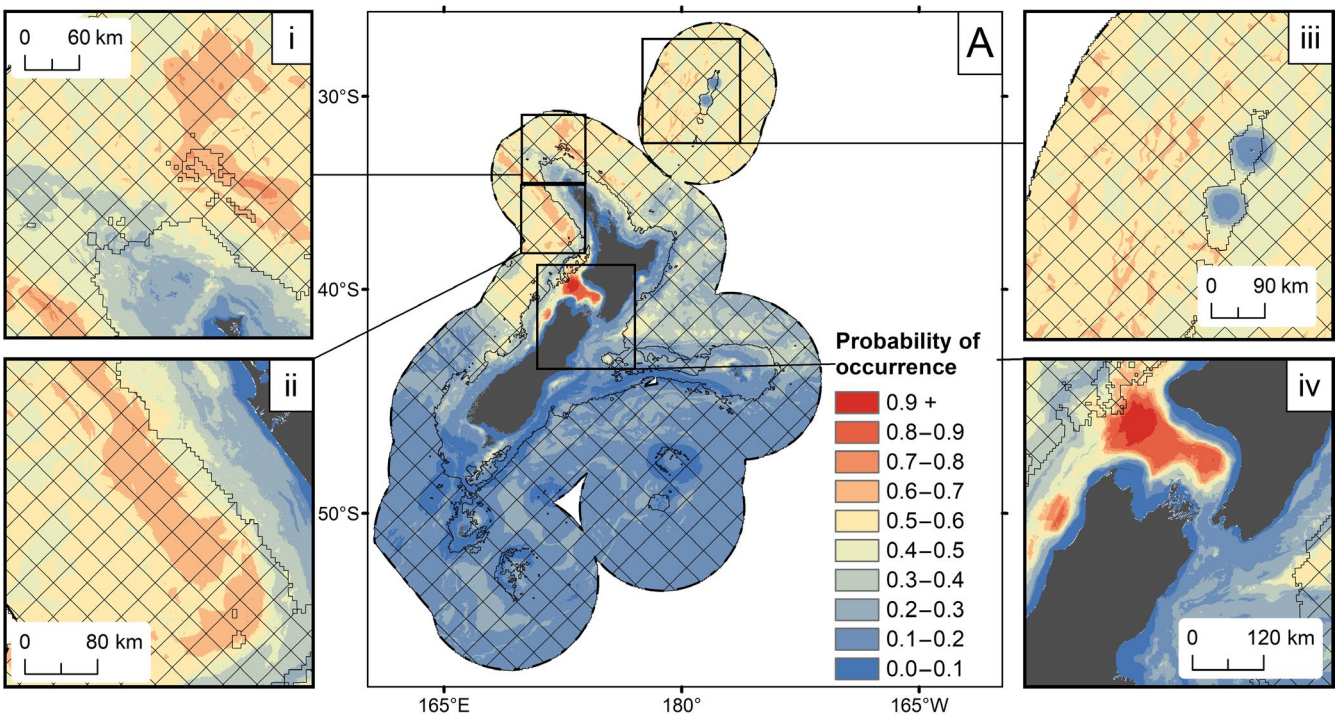

FIGURE 4 The predicted probability occurrence of blue whales (*Balaenoptera musculus musculus* and *B. m. brevicauda*) in the New Zealand EEZ modelled using bootstrapped BRTs and areas of low predicted environmental coverage depicting the lower confidence that can be placed in the predicted probability occurrence (criss-cross black line). Inset maps: (i) north of North Island including the North Cape and West Norfolk Ridge; (ii) offshore waters of west coast of North Island including a northern section of the Challenger Plateau; (iii) Kermadec islands, Lau-Colville Ridge and Kermadec Ridge; (iv) south of the North Island and north of the South Island including Tasman and Golden Bays, South Taranaki Bight and the Cook Strait. See Figure 1a, for place names

evidence required for spatial risk assessments and conservation planning. The model outputs are useful as direct inputs (e.g., spatially explicit risk assessments) or to highlight areas requiring further investigation (e.g., spatial estimates of areas and or taxa requiring further sampling). However, information gaps remain for most species, i.e., particularly for rarely sighted offshore taxa, and those with low sample size. Spatial bias in sampling effort requires careful consideration for use in robust decision making. Further, both RES and BRT models produce estimates of suitable habitat, which may be more extensive than the actual occupied habitat, particularly for those species whose ranges have been reduced through whaling or habitat degradation.

4.1 | Critical appraisal of model outputs

BRT models used to predict cetacean species' occurrence generally performed well and all models were considered useful (AUC > 0.75). Distributions of individual species' occurrence were in line with information on known distributions of some species, e.g., Māui dolphins (Derville et al., 2016), Hector's dolphins (MacKenzie & Clement, 2014), dusky dolphins (Würsig, Lynn, Jefferson, & Mullin, 1998) and for specific regional areas, e.g., the Hauraki Gulf (Dwyer et al., 2016). In addition, cetacean distributions were visually congruent with recorded sightings data (used to train the models) and NOMAD (true) presence/absence data. The independent statistical validation of models for bottlenose, common, Hector's and dusky dolphins and Bryde's and killer whales with independent true presence/absence NOMAD sightings records also showcase that these presence/relative absence models had predictive power (AUC ranging from 0.68 to 0.96), although the NOMAD data are limited to areas of interest for fisheries and therefore also suffer from spatial bias.

RES estimates for rarely sighted cetacean species were produced using a combination of expert opinion and information available from the literature (see Kaschner et al., 2006 and references therein). The RES approach is a conceptually simple modelling approach, and has been used widely (Kaschner et al., 2011; Kesner-Reyes et al., 2016), with comparable results to other presence-only models such as GARP—Genetic Algorithm for Rule Set Production, MaxEnt—Maximum Entropy Modeling, GLMs—generalised linear models and GAMs—generalised additive models (Kesner-Reyes et al., 2016). However, RES estimates rely on basic information of species' environmental preference; information which is still less certain for some of the rarest species. For the majority of species examined here, estimated probability distributions using RES were consistent with the limited sighting and coastal stranding records available for these rarely sighted species, providing evidence that the broad-scale environmental niches of individual species were effectively captured. Future iterations of RES models may benefit from the selection of a

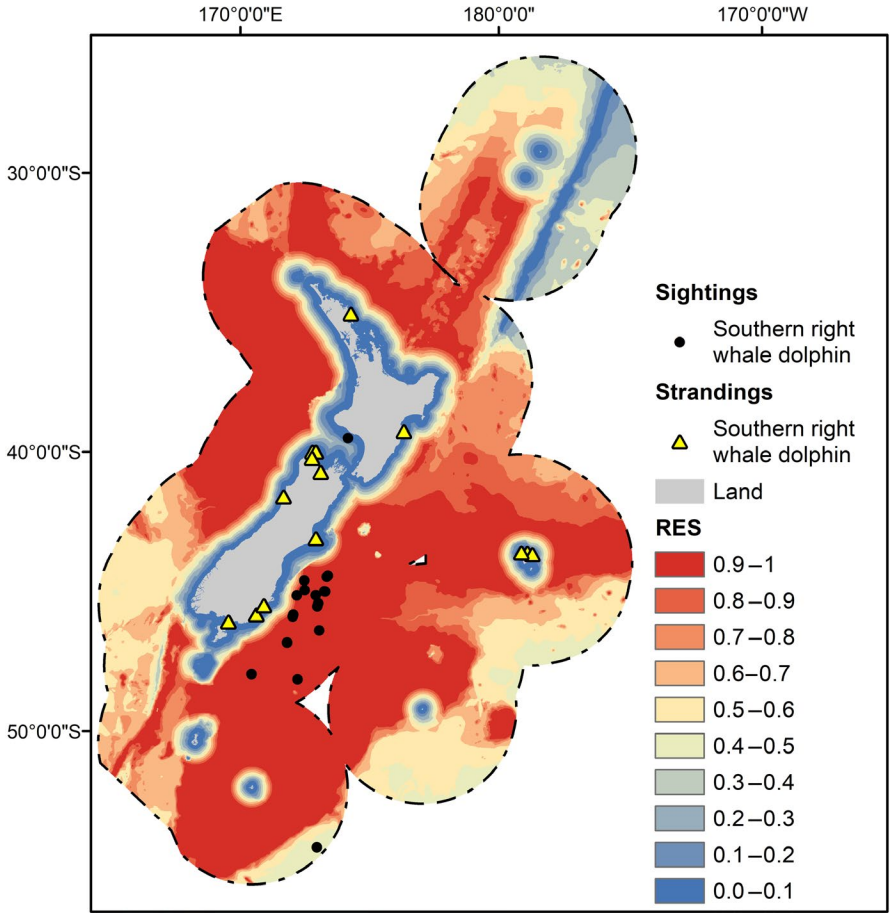

FIGURE 5 Predicted RES scores for southern right whale dolphin (*Lissodelphis peronii*), ranging from less suitable (blue) to very suitable (red). Predicted RES scores are shown with sightings at sea and location of recorded strandings (from the DOC marine mammal strandings database). Sightings at sea and location of recorded strandings were not used as inputs in the model but were used as a visual validation only

greater number of environmental predictor variables, and/or, on an individual species level, the selection of additional, and potentially more ecologically meaningful, variables. Given the low levels of information available for the species occurrences predicted using RES, these predictions remain the best available information but should be used cautiously.

4.2 | Drivers of cetacean distribution

At large spatial scales, the environmental preferences of cetaceans were primarily associated with broad-scale oceanographic conditions such as temperature and depth that typically vary on scales of hundreds of kms. For example, water temperature related predictors (SST and TempRes) were important predictors across most species. Although it is unclear whether this is an indication of the thermal tolerance/preference of these species, the preference/tolerance of their prey, or a combination of the two. Several species (e.g., common dolphin, Māui dolphin, blue whale, fin whale, bottlenose dolphin and Bryde's whales) clearly displayed preferences for northern high salinity, nutrient poor subtropical waters located to the west and north of New Zealand. For the large baleen whales these nutrient poor waters are often influenced by seasonal upwellings that stimulate patches of productivity. In particular, blue whales are associated with these upwellings near Australia and Chile (Buchan & Quiñones, 2016; Gill et al., 2011), similar to the high probability of occurrence predicted in the South Taranaki Bight here. The environmental variables that affect pelagic dolphin distribution are complex with different drivers depending on location. Small schooling fishes are the primary prey of pelagic dolphins, therefore understanding predictors of schooling fish distribution can be useful when determining the distribution of species such as dusky dolphins (e.g., Meynier et al., 2008; Selzer & Payne, 1988).

At finer spatial scales, cetacean environmental preferences were primarily associated with localised environmental conditions such as mixed layer depth, water turbidity and tidal currents that typically vary on scales or 1–10 s of km. This was particularly the case for species which frequently inhabit inshore areas where environmental gradients were steepest. For the more coastal species such as Māui and bottlenose dolphins and Bryde's whales, sufficient nearshore prey is available for year-round populations to persist (Carroll et al., 2019; Constantine, Iwata, Nieukirk, & Penry, 2018; Hartel, Constantine, & Torres, 2014; Miller, Lalas, Dawson, Ratz, & Slooten, 2013). Blue whales frequently aggregate in areas of nearshore upwelling that drive zooplankton productivity during the austral summer months from approximately January to April (Barlow et al., 2018; Torres, 2013).

Southern right whale, dusky dolphin and Hector's dolphin displayed clear large-scale preferences for cooler, low salinity, nutrient rich waters of subantarctic origin located to the south and east of the

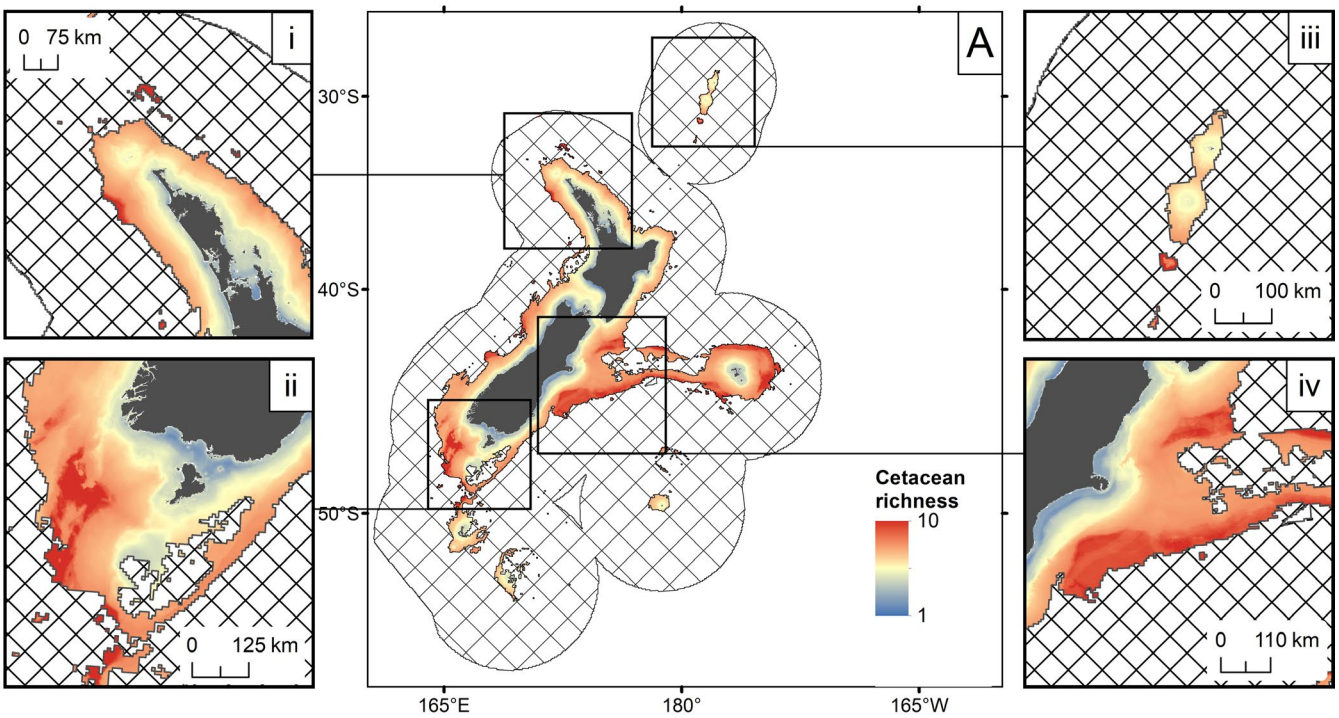

FIGURE 6 Cetacean richness estimates (using predicted occurrence derived from 15 BRT models and 15 RES models) for areas of adequate environmental coverage. Crisscrossing black lines represent areas with low environmental coverage. Inset maps: (i) north of North Island including the North Cape and Hauraki Gulf; (ii) south of the South Island including North East Island, Stewart Island and parts of Fiordland Coast; (iii) Kermadec islands and Kermadec Ridge; (iv) North East of the South Island including the western parts of the Chatham Rise

study area (e.g., primarily south of Chatham Rise). Hector's dolphin inhabit nearshore turbid waters, dispersing slightly more offshore in winter compared to summer (MacKenzie & Clement, 2014; Roberts, Webber, Roe, Edwards, & Doonan, 2019). These movements are associated with their preferred prey of red cod (*Pseudophycis bachus*), ahuru (*Auchenocerros punctatus*), sprat (*Sprattus* sp.) and javelinfish (*Lepidorhynchus denticulatus*; Miller et al., 2013). Southern right whales currently use the subantarctic Auckland Islands as a primary winter breeding ground, moving further south to feed (Childerhouse, Double, & Gales, 2010; Patenaude et al., 1998). Historically, they were more widely distributed pre-whaling (Jackson et al., 2016; Torres et al., 2013), but as the population increases, southern right whales are more frequently observed around mainland New Zealand (Carroll et al., 2014). Social factors most likely influence the whales' primary winter aggregation in Port Ross, Auckland Islands where increasing numbers of whales come into this harbour despite the availability of similar habitat on the island (Rayment et al., 2015). With predicted changes in prey availability (Torres et al., 2013) and increasing population size (Carroll et al., 2013) we suggest that southern right whales will continue to re-establish former habitat around New Zealand and the Kermadec—Louisville region to the northeast.

Offshore species and migratory species, e.g., minke, sei, sperm, killer, humpback and pilot whales, were more generalist and tended to show preference for depth related environmental variables over temperature related variables. Sperm and pilot whale preferences for deeper offshore waters around New Zealand are indicative of the location of their preferred prey, deep-water squids and mesopelagic fishes often associated with canyons and trenches (Beatson, O'Shea, & Ogle, 2007; Gaskin & Cawthorn, 1967; Giorli & Goetz, 2019; Guerra et al., 2017). Minke and sei whales' habitat use patterns are poorly known in New Zealand (Baker et al., 2019), but it is likely they are migrating through New Zealand waters between southern feeding grounds and warmer water breeding grounds. Sei whales are sighted in north-eastern New Zealand waters, with one individual frequently sighted in the Hauraki Gulf (R. Constantine, unpublished data). Humpback whales typically migrate through New Zealand waters as they move between Antarctic feeding grounds and Oceania breeding grounds (Constantine, Russell, Gibbs, Childerhouse, & Baker, 2007; Dawbin, 1960), stopping briefly at the northern Kermadec Islands in social aggregations (Riekkola et al., 2018) and near Fiordland where some foraging occurs before they complete their southern migration to Antarctic waters (Gales et al., 2009). Killer whales range widely throughout all New Zealand offshore and nearshore waters feeding on a variety of benthic and demersal prey (Visser, 2000), typical of the species' global distribution and habitat use patterns (Ford, 2018).

Some species with widely distributed recorded locations had poorer model fits than species with restricted ranges, perhaps reflecting the cosmopolitan distribution of the former (e.g., moderate explained deviance and AUC scores for killer whale and bottlenose dolphin) and the more aggregated nature of others for the latter (e.g., high explained deviance and AUC scores for Māui dolphin, Hector's dolphin). Evidence from previous studies have indicated that species

with limited geographic ranges and/or environmental tolerances are generally better modelled than those with greater ranges (Morán-Ordóñez, Lahoz-Monfort, Elith, & Wintle, 2017; Stephenson et al., 2018; Thomson et al., 2014) because widespread species are less likely to have sharp easily identifiable environmental thresholds that clearly delineate their environmental niche (Morán-Ordóñez et al., 2017). For species with limited ranges, the best model fits were commonly located closer to shore where sampling effort was highest.

Additionally, reduced model fit could be influenced by historical events, human activities, population and species dynamics (e.g., migration, competition, predation, and for many cetacean species, social interactions; Elith & Leathwick, 2009) and temporal environmental patterns (e.g., diurnal, tidal, seasonal and annual patterns; fluctuating weather patterns; and prey distributions) which were not accounted for here. Despite these factors not being considered in a quantitative manner, model outputs are still valid for management purposes, but it should be noted that the representation of species' probability of occurrence are a smoothed representation of the raw data (spatially and temporally; Stephenson et al., 2018). Both RES and BRT analyses are correlative models and, in many cases, rely on biotic processes such as predation to be represented by environmental variables as proxies. Prey distribution is likely to be of particular importance for the accurate prediction of cetacean distributions (Bluhm et al., 2007; Etnoyer et al., 2006; Tynan et al., 2005). Further work on integrating predictors not accounted for here, for example by including prey species in a joint species distribution model (hierarchical probit regressions with latent factors as in Ovaskainen et al., 2017) or population dynamics and social interactions in mechanistically focussed models such as process-based range modeling (Dormann et al., 2012; Evans, Merow, Record, McMahon, & Enquist, 2016) may provide further insight into important drivers of cetacean distributions not accounted for here. As a first step, defining prey target species and producing prey distributions for each cetacean taxa would be of interest for future distribution modelling, but we note that this is particularly challenging as it requires an understanding of predator-prey relationships in time and space over the vast New Zealand EEZ.

A relatively simple representation of cetacean richness was provided. This showcased the high cetacean richness across extensive areas within New Zealand, particularly offshore. Hotspots of cetacean richness in offshore areas (Macquarie Ridge, the western edge of the Bounty Trough and along the southern and northern edges of the Chatham Rise) are known to have high biological productivity and are likely to be important foraging areas for several cetacean species. Although inshore areas generally had lower richness estimates, these remain important from a conservation point of view for many of the species with limited ranges (e.g., Maui and Hector's dolphin). Indeed, the distribution of species richness may be biased towards areas where species with large overlapping ranges occur, i.e., those species with restricted non-overlapping ranges will not contribute greatly towards the summed species richness (Veach, Minin, Pouzols, & Moilanen, 2017). In order to account for range size bias, systematic conservation planning software (e.g., Moilanen, Kujala, & Possingham, 2009) could be used to identify important conservations areas (i.e., balance out selection of areas with high richness whilst maintaining adequate species representativeness). Further, richness estimates from stacked species distribution models could be improved by incorporating estimates from a macroecological model (as described in Calabrese et al., 2014) as well as including estimates of model uncertainty (Moilanen et al., 2006). Further work is currently underway to investigate these important issues.

4.3 | Use of at-sea cetacean sightings records for species distribution modelling

Dedicated cetacean census surveys do not exist in New Zealand due to logistic and financial constraints. The majority of data available are for nearshore cetaceans, causing large biases in the availability of data to parameterise species distribution models, with typically fewer data records outside highly populated areas or areas of intensive tourism, resulting in limitations of model predictions outside areas where sightings data occurred (Derville et al., 2018). For example, Bryde's whale predicted occurrence was predominately constrained to the Hauraki Gulf due to the high number of sightings in this area (year-round population of Bryde's whales that are the focus of a whale watch industry and research programmes), despite knowledge that these are also an offshore species in New Zealand and elsewhere (Constantine et al., 2018). Recent surveys (data not included in our models) highlight that the Kaikoura coast along the north east of the South Island also have high occurrences of blue whales (Barlow et al., 2018) although this is not apparent from the models presented here due to the low number of sightings available for this location in our study. With the inclusion of newly acquired data, or the inclusion of mechanistic relationships in the distribution models, prediction accuracy and usefulness is likely to improve.

In this study, the final dataset contained over 14,000 cetacean sightings records from multiple sources, which were unequally distributed across both the study area and among species. Ideally, true presence/absence with known effort (such as in NOMAD database and Marine Mammal Observer data from seismic surveys) would be used for all future models; however, these data also suffer from spatial bias and are currently only available for a limited number of species (e.g., see Roberts et al., 2019). Some species are vulnerable to fisheries interactions and therefore may be subject to targeted fishery observer campaigns to determine overlap of species with fishing activity (e.g., NOMAD data). Hector's and Māui dolphins, both coastal species, are the only species that are the subject of large-scale aerial surveys (e.g., MacKenzie & Clement, 2014). With large whale species recovering from commercial whaling, public sightings records have increased for some whales when in nearshore waters; this has been valuable to understanding their shifts in distribution e.g., southern right whales (Carroll et al., 2014). Future data from dedicated research and public sightings will be valuable in understanding any shifts in species' range due to climate change effects on environmental variables that drive distribution,

as seen elsewhere (e.g., Gulf of St Lawrence, Ramp, Delarue, Palsbøll, Sears, & Hammond, 2015).

The spatial sample bias in this study was somewhat addressed by the weighting of records and subsampling of non-random absence/background data (e.g., relative absences) and through bootstrapping of the BRT models. Future model iterations could explore the effects of spatial autocorrelation (i.e., the tendency for clustering of records spatially) through inclusion of a spatial autocorrelation term in the model (e.g., calculation of a residual autocovariate [RAC] variable for each species as in Crase, Liedloff, & Wintle, 2012), or weighting of species presence based on environmental coverage (e.g., as in Smith et al., 2013; Valavi, Elith, Lahoz-Monfort, & Guillera-Arroita, 2018).

The use of relative absences (i.e., the presence of another species for which the model is not being predicted was used as evidence that the species of interest was absent) has been shown to improve model accuracy when compared to the use of random background points (pseudo absences; Phillips et al., 2009). However, differences in detectability varies substantially between species, especially for offshore species, deep diving species and migratory species that are less available to be sighted (Barlow, Gerodette, & Forcada, 2001; Würsig et al., 1998). The influence that these differences in detectability (also sometimes called "probability of sampling" or "catch rate" for other marine taxa) is a parameter that is difficult to estimate and the influence that this has on model prediction is unknown (Anderson et al., 2016). However, simulation and virtual ecologist approaches may provide possible avenues for further understanding the effect that species detectability may have on model prediction (Meynard, Leroy, & Kaplan, 2019).

5 | CONCLUSION

BRT models used to predict cetacean species' occurrence generally performed well with the distributions of individual species in line with the limited information on their known distributions. For a small subset of species, the usefulness of the BRT models was further validated using an independent set of true presence/absence data. The simpler RES model estimates for rarely sighted cetacean species rely on basic information of species' environmental preference. Given the low levels of information available for these species, RES remains the best available information. However, information gaps remain for many species (i.e., particularly for the less frequently observed offshore taxa with fewer than 50 recorded sightings). Further information on important drivers of cetacean distribution and how these differ between species indicates several offshore areas which are known to have high biological productivity as potentially important foraging areas for several cetacean species. Although inshore areas generally had lower richness estimates these remain important from a conservation point of view for many of the species with limited ranges (e.g., Maui and Hector's dolphin). Issues relating to sample size and spatial sampling bias require careful consideration for use in robust decision making. Here we have addressed some of these potential pitfalls for decision making by providing two spatially explicit estimates of model uncertainty: estimates of individual species' uncertainty of the distributions (through bootstrapping of the BRT models) and estimates of the coverage of the environmental space from all species recorded sightings. Predicted spatial distributions, their associated uncertainty layers and summaries of these layers highlighting important areas (i.e., cetacean richness estimates) are a significant step towards providing the evidence necessary as a baseline for spatial risk assessments and conservation planning.

ACKNOWLEDGEMENTS

This work was funded by the New Zealand Ministry of Primary Industries (project PR2O1401) and the NIWA Coasts and Oceans Programme (Project: COME1903). We thank J. Leathwick for discussing and providing useful insights on modelling species distributions, M. Pinkerton for providing many of the environmental predictors used in this study, H. Hendriks for helping with the DOC data, M. Cawthorn for his database of cetacean sightings, L Torres for advice and the initial funding for this research and A. van Helden, D. Clement and K. Hupman for their advice and expert opinion on cetacean distributions.

DATA AVAILABILITY STATEMENT

Cetacean sightings and distributions generated for this study will be made available from the MPI managed NABIS website (https://maps.mpi.govt.nz).

ORCID

Fabrice Stephenson https://orcid.org/0000-0002-9500-5204

REFERENCES

Anderson, O. F., Guinotte, J. M., Rowden, A. A., Tracey, D. M., Mackay, K. A., & Clark, M. R. (2016). Habitat suitability models for predicting the occurrence of vulnerable marine ecosystems in the seas around New Zealand. *Deep Sea Research Part I: Oceanographic Research Papers*, *115*, 265–292. https://doi.org/10.1016/j.dsr.2016.07.006

Anderson, R. P. (2003). Real vs. artefactual absences in species distributions: Tests for *Oryzomys albigularis* (Rodentia: Muridae) in Venezuela. *Journal of Biogeography*, *30*, 591–605. https://doi.org/10.1046/j.1365-2699.2003.00867.x

Baker, C. S., Boren, L., Childerhouse, S., Constantine, R., van Helden, A., Lundquist, D., … Rolfe, J. R. (2019). *Conservation status of New Zealand marine mammals, 2019. New Zealand Threat Classification Series 29*. Wellington, New Zealand: Department of Conservation.

Barlow, D. R., Torres, L. G., Hodge, K. B., Steel, D., Baker, C. S., Chandler, T. E., … Klinck, H. (2018). Documentation of a New Zealand blue whale population based on multiple lines of evidence. *Endangered Species Research*, *36*, 27–40. https://doi.org/10.3354/esr00891

Barlow, J., Gerodette, T., & Forcada, J. (2001). Factors affecting perpendicular sighting distances on shipboard line-transect surveys for cetaceans. *Journal of Cetacean Research and Management*, *3*, 201–212.

Beatson, E., O'Shea, S., & Ogle, M. (2007). First report on the stomach contents of long-finned pilot whales, *Globicephala melas*, stranded in New Zealand. *New Zealand Journal of Zoology*, *34*, 51–56.

Behrenfeld, M. J., & Falkowski, P. G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography*, *42*, 1–20. https://doi.org/10.4319/lo.1997.42.1.0001

Bluhm, B., Coyle, K., Konar, B., & Highsmith, R. (2007). High gray whale relative abundances associated with an oceanographic front in the south-central Chukchi Sea. *Deep Sea Research Part II: Topical Studies in Oceanography*, *54*, 2919–2933. https://doi.org/10.1016/j.dsr2.2007.08.015

Bradford-Grieve, J. M., Lewis, K. B., & Stanton, B. R. (1991). Advances in New Zealand oceanography, 1967–91. *New Zealand Journal of Marine and Freshwater Research*, *25*, 429–441. https://doi.org/10.1080/00288330.1991.9516495

Bradford-Grieve, J., Probert, K., Lewis, K., Sutton, P., Zeldis, J., & Orpin, A. (2006). *New Zealand shelf region*. Cambridge, MA: Harvard University Press.

Bräger, S., & Schneider, K. (1998). Near-shore distribution and abundance of dolphins along the West Coast of the South Island, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, *32*, 105–112. https://doi.org/10.1080/00288330.1998.9516809

Buchan, S. J., & Quiñones, R. A. (2016). First insights into the oceanographic characteristics of a blue whale feeding ground in northern Patagonia, Chile. *Marine Ecology Progress Series*, *554*, 183–199. https://doi.org/10.3354/meps11762

Calabrese, J. M., Certain, G., Kraan, C., & Dormann, C. F. (2014). Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, *23*, 99–112. https://doi.org/10.1111/geb.12102

Carroll, E. L., Childerhouse, S. J., Fewster, R. M., Patenaude, N. J., Steel, D., Dunshea, G., … Baker, C. S. (2013). Accounting for female reproductive cycles in a superpopulation capture–recapture framework. *Ecological Applications*, *23*, 1677–1690. https://doi.org/10.1890/12-1657.1

Carroll, E. L., Gallego, R., Sewell, M. A., Zeldis, J., Ranjard, L., Ross, H. A., … Constantine, R. (2019). Multi-locus DNA metabarcoding of zooplankton communities and scat reveal trophic interactions of a generalist predator. *Scientific Reports*, *9*, 281. https://doi.org/10.1038/s41598-018-36478-x

Carroll, E. L., Rayment, W. J., Alexander, A. M., Baker, C. S., Patenaude, N. J., Steel, D., … Childerhouse, S. (2014). Reestablishment of former wintering grounds by New Zealand southern right whales. *Marine Mammal Science*, *30*, 206–220. https://doi.org/10.1111/mms.12031

Childerhouse, S., Double, M., & Gales, N. (2010). *Satellite tracking of southern right whales (Eubalaena australis) at the Auckland Islands, New Zealand*. Unpublished report (SC/62/BRG19) presented to the Scientific Committee of the International Whaling Commission, Cambridge, UK.

Compton, T. J., Morrison, M. A., Leathwick, J. R., & Carbines, G. D. (2012). Ontogenetic habitat associations of a demersal fish species, *Pagrus auratus*, identified using boosted regression trees. *Marine Ecology Progress Series*, *462*, 219–230. https://doi.org/10.3354/meps09790

Constantine, R., Iwata, T., Nieukirk, S., & Penry, G. S. (2018). Future directions in research on Bryde's whales. *Frontiers in Marine Science*, *5*, 1–7. https://doi.org/10.3389/fmars.2018.00333

Constantine, R., Russell, K., Gibbs, N., Childerhouse, S., & Baker, C. S. (2007). Photo-identification of humpback whales (*Megaptera novaeangliae*) in New Zealand waters and their migratory connections to breeding grounds of Oceania. *Marine Mammal Science*, *23*, 715–720. https://doi.org/10.1111/j.1748-7692.2007.00124.x

Crase, B., Liedloff, A. C., & Wintle, B. A. (2012). A new method for dealing with residual spatial autocorrelation in species distribution models. *Ecography*, *35*, 879–888. https://doi.org/10.1111/j.1600-0587.2011.07138.x

Davidson, A. D., Boyer, A. G., Kim, H., Pompa-Mansilla, S., Hamilton, M. J., Costa, D. P., … Brown, J. H. (2012). Drivers and hotspots of extinction risk in marine mammals. *Proceedings of the National Academy of Sciences of the United States of America*, *109*, 3395–3400. https://doi.org/10.1073/pnas.1121469109

Dawbin, W. H. (1960). An analysis of the New Zealand catches of Humpbank Whales from 1947 to 1958. *Norsk Hvalfangst Tidende*, *2*, 61–75.

Derville, S., Constantine, R., Baker, C. S., Oremus, M., & Torres, L. G. (2016). Environmental correlates of nearshore habitat distribution by the Critically Endangered Māui dolphin. *Marine Ecology Progress Series*, *551*, 261–275.

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*, 1657–1673. https://doi.org/10.1111/ddi.12782

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., … Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*, 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Dormann, C. F., Schymanski, S. J., Cabral, J., Chuine, I., Graham, C., Hartig, F., … Singer, A. (2012). Correlation and process in species distribution models: Bridging a dichotomy. *Journal of Biogeography*, *39*, 2119–2131. https://doi.org/10.1111/j.1365-2699.2011.02659.x

Duffy, C. A., Baker, C. S., & Constantine, R. (2015). Observation and identification of marine mammals during two recent expeditions to the Kermadec Islands, New Zealand. *Bulletin of the Auckland Museum*, *20*, 501–510.

Dwyer, S., Clement, D., Pawley, M., & Stockin, K. (2016). Distribution and relative density of cetaceans in the Hauraki Gulf, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, *50*, 457–480. https://doi.org/10.1080/00288330.2016.1160942

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., … Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, *29*, 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

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*, 677–697. https://doi.org/10.1146/annurev.ecolsys.110308.120159

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*, 802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Etnoyer, P., Canny, D., Mate, B. R., Morgan, L. E., Ortega-Ortiz, J. G., & Nichols, W. J. (2006). Sea-surface temperature gradients across blue whale and sea turtle foraging trajectories off the Baja California Peninsula, Mexico. *Deep Sea Research Part II: Topical Studies in Oceanography*, *53*, 340–358. https://doi.org/10.1016/j.dsr2.2006.01.010

Evans, M. E., Merow, C., Record, S., McMahon, S. M., & Enquist, B. J. (2016). Towards process-based range modeling of many species. *Trends in Ecology & Evolution*, *31*, 860–871. https://doi.org/10.1016/j.tree.2016.08.005

Ferrier, S., & Guisan, A. (2006). Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, *43*, 393–404. https://doi.org/10.1111/j.1365-2664.2006.01149.x

Ford, J. (2018). Killer whale. In B. Würsig, J. G. M. Thewissen, & K. M. Kovacs (Eds.), *Encyclopedia of marine mammals* (3rd ed.). London, UK: Academic Press.

Friedman, J. H., & Meulman, J. J. (2003). Multiple additive regression trees with application in epidemiology. *Statistics in Medicine*, *22*, 1365–1381. https://doi.org/10.1002/sim.1501

Gales, N., Double, M. C., Robinson, S., Jenner, C., Jenner, M., King, E., … Raymond, B. (2009). *Satellite tracking of southbound East Australian humpback whales* (Megaptera novaeangliae)*: Challenging the feast or famine model for migrating whales*. Unpublished report (SC/61/SH17) presented to the Scientific Committee of the International Whaling Commission, Cambridge, UK.

Gaskin, D. E., & Cawthorn, M. W. (1967). Diet and feeding habits of the sperm whale (*Physeter catodon* L.) in the cook strait region of New Zealand. *New Zealand Journal of Marine and Freshwater Research*, *1*, 156–179.

Gill, P. C., Morrice, M. G., Page, B., Pirzl, R., Levings, A. H., & Coyne, M. (2011). Blue whale habitat selection and within-season distribution in a regional upwelling system off southern Australia. *Marine Ecology Progress Series*, *421*, 243–263. https://doi.org/10.3354/meps08914

Giorli, G., & Goetz, K. T. (2019). Foraging activity of sperm whales (*Physeter macrocephalus*) off the east coast of New Zealand. *Scientific Reports*, *9*, 12182. https://doi.org/10.1038/s41598-019-48417-5

Gorman, R. M., Bryan, K. R., & Laing, A. K. (2003). Wave hindcast for the New Zealand region: Deep-water wave climate. *New Zealand Journal of Marine and Freshwater Research*, *37*, 589–612. https://doi.org/10.1080/00288330.2003.9517191

Guerra, M., Hickmott, L., van der Hoop, J., Rayment, W., Leunissen, E., Slooten, E., & Moore, M. (2017). Diverse foraging strategies by a marine top predator: Sperm whales exploit pelagic and demersal habitats in the Kaikōura submarine canyon. *Deep Sea Research Part I: Oceanographic Research Papers*, *128*, 98–108. https://doi.org/10.1016/j.dsr.2017.08.012

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. *Ecology Letters*, *8*, 993–1009. https://doi.org/10.1111/j.1461-0248.2005.00792.x

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., … Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, *16*, 1424–1435. https://doi.org/10.1111/ele.12189

Guisan, A., & Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, *135*, 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Hamner, R. M., Pichler, F. B., Heimeier, D., Constantine, R., & Baker, C. S. (2012). Genetic differentiation and limited gene flow among fragmented populations of New Zealand endemic Hector's and Maui's dolphins. *Conservation Genetics*, *13*, 987–1002. https://doi.org/10.1007/s10592-012-0347-9

Hartel, E. F., Constantine, R., & Torres, L. G. (2014). Changes in habitat use patterns by bottlenose dolphins over a 10-year period render static management boundaries ineffective. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *25*, 538–547.

Hijmans, R. J., Phillips, S., Leathwick, J., & Elith, J. (2017). *dismo: Species Distribution Modeling R package version 1.1-4*. Retrieved from https://CRAN.R-project.org/package=dismo

Hijmans, R. J., & van Etten, J. (2012). *raster: Geographic analysis and modeling with raster data. R package version 2.0-12*. Retrieved from http://CRAN.R-project.org/package=raster

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C., & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, *199*, 142–152. https://doi.org/10.1016/j.ecolmodel.2006.05.017

IUCN (2001). *IUCN Red List categories and criteria*. Gland, Switzerland and Cambridge, UK: IUCN Species Survival Commission.

IUCN (2019). *The IUCN Red List of Threatened Species. [Online]. Version: 2019-1*. Retrieved from https://www.iucnredlist.org

Jackson, J. A., Carroll, E. L., Smith, T. D., Zerbini, A. N., Patenaude, N. J., & Baker, C. S. (2016). An integrated approach to historical population assessment of the great whales: Case of the New Zealand southern right whale. *Royal Society Open Science*, *3*, 150669. https://doi.org/10.1098/rsos.150669

Jepson, P. D., Deaville, R., Barber, J. L., Aguilar, À., Borrell, A., Murphy, S., … Law, R. J. (2016). PCB pollution continues to impact populations of orcas and other dolphins in European waters. *Scientific Reports*, *6*, 18573. https://doi.org/10.1038/srep18573

Kaschner, K., Tittensor, D. P., Ready, J., Gerrodette, T., & Worm, B. (2011). Current and future patterns of global marine mammal biodiversity. *PLoS ONE*, *6*, e19653. https://doi.org/10.1371/journal.pone.0019653

Kaschner, K., Watson, R., Trites, A. W., & Pauly, D. (2006). Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. *Marine Ecology Progress Series*, *316*, 285–310. https://doi.org/10.3354/meps316285

Kesner-Reyes, K., Kaschner, K., Kullander, S., Garilao, C., Barile, J., & Froese, R. (2016). *AquaMaps: algorithm and data sources for aquatic organisms. Version (04/2012)*. Retrieved from www.fishbase.org

Leathwick, J., Elith, J., Francis, M., Hastie, T., & Taylor, P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: An analysis using boosted regression trees. *Marine Ecology Progress Series*, *321*, 267–281. https://doi.org/10.3354/meps321267

Leathwick, J., Rowden, A., Nodder, S., Gorman, R., Bardsley, S., Pinkerton, M., … Goh, A. (2012). A Benthic-optimised marine environment classification (BOMEC) for New Zealand waters. In *New Zealand Aquatic Environment and Biodiversity Report*. Wellington, New Zealand: Ministry of Fisheries.

MacKenzie, D. L., & Clement, D. M. (2014). *Abundance and distribution of ECSI Hector's dolphins–supplementary material. Supplement to the New Zealand Aquatic and Environment and Biodiversity Report*. Ministry for Primary Industries, Wellington, New Zealand.

Meynard, C. N., Leroy, B., & Kaplan, D. M. (2019). Testing methods in species distribution modelling using virtual species: What have we learnt and what are we missing? *Ecography*, *42*, 2021–2036. https://doi.org/10.1111/ecog.04385

Meynier, L., Pusineri, C., Spitz, J., Santos, M. B., Pierce, G. J., & Ridoux, V. (2008). Intraspecific dietary variation in the short-beaked common dolphin *Delphinus delphis* in the Bay of Biscay: Importance of fat fish. *Marine Ecology Progress Series*, *354*, 277–287. https://doi.org/10.3354/meps07246

Miller, B. S., Collins, K., Barlow, J., Calderan, S., Leaper, R., Mcdonald, M., … Double, M. C. (2014). Blue whale vocalizations recorded around New Zealand: 1964–2013. *The Journal of the Acoustical Society of America*, *135*, 1616–1623. https://doi.org/10.1121/1.4863647

Miller, E., Lalas, C., Dawson, S., Ratz, H., & Slooten, E. (2013). Hector's dolphin diet: The species, sizes and relative importance of prey eaten by *Cephalorhynchus hectori*, investigated using stomach content analysis. *Marine Mammal Science*, *29*, 606–628.

Mitchell, J. S., Mackay, K. A., Neil, H. L., Mackay, E. J., Pallentin, A., & Notman, P. (2012). *Undersea New Zealand, 1:5,000,000. NIWA Chart, Miscellaneous Series No. 92*.

Moilanen, A., Kujala, H., & Possingham, H. P. (2009). *The Zonation framework and software for conservation prioritization*. Oxford, UK: Oxford University Press.

Moilanen, A., Runge, M. C., Elith, J., Tyre, A., Carmel, Y., Fegraus, E., … Ben-Haim, Y. (2006). Planning for robust reserve networks using uncertainty analysis. *Ecological Modelling*, *199*, 115–124. https://doi.org/10.1016/j.ecolmodel.2006.07.004

Morán-Ordóñez, A., Lahoz-Monfort, J. J., Elith, J., & Wintle, B. A. (2017). Evaluating 318 continental-scale species distribution models over a 60-year prediction horizon: What factors influence the reliability of predictions? *Global Ecology and Biogeography*, *26*, 371–384. https://doi.org/10.1111/geb.12545

Oremus, M., Gales, R., Dalebout, M. L., Funahashi, N., Endo, T., Kage, T., … Baker, S. C. (2009). Worldwide mitochondrial DNA diversity and phylogeography of pilot whales (*Globicephala* spp.). *Biological Journal of the Linnean Society*, *98*, 729–744. https://doi.org/10.1111/j.1095-8312.2009.01325.x

Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., … Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, *20*, 561–576. https://doi.org/10.1111/ele.12757

Patenaude, N. J., Baker, C. S., & Gales, N. J. (1998). Observations of southern right whales on New Zealand's subantarctic wintering grounds. *Marine Mammal Science*, *14*, 350–355. https://doi.org/10.1111/j.1748-7692.1998.tb00726.x

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J., & Ferrier, S. (2009). Sample selection bias and presence-only distribution models: Implications for background and pseudo-absence data. *Ecological Applications*, *19*, 181–197. https://doi.org/10.1890/07-2153.1

Pinkerton, M., Gall, M., Wood, S., & Zeldis, J. (2018). Measuring the effects of bivalve mariculture on water quality in northern New Zealand using 15 years of MODIS-Aqua satellite observations. *Aquaculture Environment Interactions*, *10*, 529–545. https://doi.org/10.3354/aei00288

Pirotta, E., Booth, C. G., Costa, D. P., Fleishman, E., Kraus, S. D., Lusseau, D., … Harwood, J. (2018). Understanding the population consequences of disturbance. *Ecology and Evolution*, *8*, 9934–9946. https://doi.org/10.1002/ece3.4458

Pompa, S., Ehrlich, P. R., & Ceballos, G. (2011). Global distribution and conservation of marine mammals. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 13600–13605. https://doi.org/10.1073/pnas.1101525108

Porter, L., & Lai, H. Y. (2017). Marine mammals in Asian societies; trends in consumption, bait, and traditional use. *Frontiers in Marine Science*, *4*, 47. https://doi.org/10.3389/fmars.2017.00047

Pyenson, N. D. (2011). The high fidelity of the cetacean stranding record: insights into measuring diversity by integrating taphonomy and macroecology. *Proceedings of the Royal Society B: Biological Sciences*, *278*, 3608–3616.

Ramp, C., Delarue, J., Palsbøll, P. J., Sears, R., & Hammond, P. S. (2015). Adapting to a Warmer Ocean—Seasonal Shift of Baleen Whale Movements over Three Decades. *PLOS ONE*, *10*, e0121374.

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rayment, W., Dawson, S., & Webster, T. (2015). Breeding status affects fine-scale habitat selection of southern right whales on their wintering grounds. *Journal of Biogeography*, *42*, 463–474. https://doi.org/10.1111/jbi.12443

Read, A. J. (2008). The looming crisis: Interactions between marine mammals and fisheries. *Journal of Mammalogy*, *89*, 541–548. https://doi.org/10.1644/07-MAMM-S-315R1.1

Ridgeway, G. (2007). *Generalized Boosted Models: A guide to the gbm package*.

Riekkola, L., Zerbini, A. N., Andrews, O., Andrews-Goff, V., Baker, C. S., Chandler, D., … Constantine, R. (2018). Application of a multi-disciplinary approach to reveal population structure and Southern Ocean feeding grounds of humpback whales. *Ecological Indicators*, *89*, 455–465. https://doi.org/10.1016/j.ecolind.2018.02.030

Roberts, J., Webber, D. N., Roe, W. D., Edwards, C. T. T., & Doonan, I. J. (2019). *Spatial risk assessment of threats to Hector's and Māui Dolphins (*Cephalorhynchus hectori*)*. Wellington, New Zealand: New Zealand Aquatic Environment and Biodiversity Report No. 214.

Selzer, L. A., & Payne, P. M. (1988). The distribution of white-sided (*Lagenorhynchus acutus*) and common dolphins (*Delphinus delphis*) vs environmental features of the continental shelf of the northeastern United States. *Marine Mammal Science*, *4*, 141–153. https://doi.org/10.1111/j.1748-7692.1988.tb00194.x

Smith, A. N., Duffy, C., Anthony, J., & Leathwick, J. R. (2013). *Predicting the distribution and relative abundance of fishes on shallow subtidal reefs around New Zealand*. Wellington, New Zealand: Department of Conservation.

Stephenson, F., Leathwick, J. R., Geange, S. W., Bulmer, R. H., Hewitt, J. E., Anderson, O. F., … Lundquist, C. J. (2018). Using Gradient Forests to summarize patterns in species turnover across large spatial scales and inform conservation planning. *Diversity and Distributions*, *24*, 1641–1656. https://doi.org/10.1111/ddi.12787

Stevens, C. L., O'Callaghan, J. M., Chiswell, S. M., & Hadfield, M. G. (2019). Physical oceanography of New Zealand/Aotearoa shelf seas–a review. *New Zealand Journal of Marine and Freshwater Research*, 1–40. https://doi.org/10.1080/00288330.2019.1588746

Thomson, R. J., Hill, N. A., Leaper, R., Ellis, N., Pitcher, C. R., Barrett, N. S., … Edgar, J. G. (2014). Congruence in demersal fish, macroinvertebrate, and macroalgal community turnover on shallow temperate reefs. *Ecological Applications*, *24*, 287–299. https://doi.org/10.1890/12-1549.1

Thompson, K. F., Millar, C. D., Scott Baker, C., Dalebout, M., Steel, D., van Helden, A. L., & Constantine, R. (2013). A novel conservation approach provides insights into the management of rare cetaceans. *Biological Conservation*, *157*, 331–340.

Torres, L. G. (2013). Evidence for an unrecognised blue whale foraging ground in New Zealand. *New Zealand Journal of Marine and Freshwater Research*, *47*, 235–248. https://doi.org/10.1080/00288330.2013.773919

Torres, L. G., Smith, T. D., Sutton, P., MacDiarmid, A., Bannister, J., & Miyashita, T. (2013). From exploitation to conservation: Habitat models using whaling data predict distribution patterns and threat exposure of an endangered whale. *Diversity and Distributions*, *19*, 1138–1152. https://doi.org/10.1111/ddi.12069

Tyack, P. L., Zimmer, W. M. X., Moretti, D., Southall, B. L., Claridge, D. E., Durban, J. W., … Boyd, I. L. (2011). Beaked whales respond to simulated and actual navy sonar. *PLoS ONE*, *6*, e17009. https://doi.org/10.1371/journal.pone.0017009

Tynan, C. T., Ainley, D. G., Barth, J. A., Cowles, T. J., Pierce, S. D., & Spear, L. B. (2005). Cetacean distributions relative to ocean processes in the northern California Current System. *Deep Sea Research Part II: Topical Studies in Oceanography*, *52*, 145–167. https://doi.org/10.1016/j.dsr2.2004.09.024

Valavi, R., Elith, J., Lahoz-Monfort, J. J., & Guillera-Arroita, G. (2018). blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *bioRxiv*, 357798.

Veach, V., Di Minin, E., Pouzols, F. M., & Moilanen, A. (2017). Species richness as criterion for global conservation area placement leads to large losses in coverage of biodiversity. *Diversity and Distributions*, *23*, 715–726. https://doi.org/10.1111/ddi.12571

Visser, I. (2000). *Orca (*Orcinus orca*) in New Zealand waters*. Doctoral thesis, University of Auckland.

Walters, R. A., Goring, D. G., & Bell, R. G. (2001). Ocean tides around New Zealand. *New Zealand Journal of Marine and Freshwater Research*, *35*, 567–579. https://doi.org/10.1080/00288330.2001.9517023

Würsig, B., Lynn, S. K., Jefferson, T. A., & Mullin, K. D. (1998). Survey ships and aircraft. *Aquatic Mammals*, *24*, 41–50.

BIOSKETCH

Fabrice Stephenson is a quantitative ecologist with interests in explaining and modelling spatial patterns of biodiversity of marine species at various scales for use in systematic conservation planning and risk assessment.

Author contribution: F.S. K.G., B.S., C.L and A.M. conceived the ideas; K.G., J.R., A.M., F.S. and B.S collated the data; F.S., T.M. and F.B. analysed the data; F.S., K.G. and R.C. led the writing, all authors contributed to the writing.

SUPPORTING INFORMATION

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

How to cite this article: Stephenson F, Goetz K, Sharp BR, et al. Modelling the spatial distribution of cetaceans in New Zealand waters. *Divers Distrib*. 2020;26:495–516. https://doi.org/10.1111/ddi.13035