

D
R
A
F
T



Exploring the drivers of spatial distributions of basking shark (*Cetorhinus maximus*) in the South Pacific

Prepared for Department of Conservation (POP2020-03)

November 2020

Prepared by:
 Brittany Finucci
 Fabrice Stephenson
 Grady Petersen
 Malcolm P. Francis
 Matthew H. Pinkerton

For any information regarding this report please contact:

Brittany Finucci
 Fisheries Scientist
 Marine Megafauna and Fish Biology
 +64-4-386 0583
 brit.finucci@niwa.co.nz

National Institute
 of Water & Atmospheric
 Research Ltd (NIWA)

301 Evans Bay Parade
 Hataitai
 Wellington 6021
 Private Bag 14901
 Kilbirnie
 Wellington 6241

Phone +64 4 386 0300

NIWA CLIENT REPORT No:
 Report date: November 2020
 NIWA Project:

Revision	Description	Date
Version 1.0		Day Month Year
Version 1.1	Amendments to ...	Day Month Year

Sign, manually or electronically, in the first column. Write or type your name in the third column

Quality Assurance Statement		
	Reviewed by:	Tom Brough

© All rights reserved. This publication may not be reproduced or copied in any form without the permission of the copyright owner(s). Such permission is only to be given in accordance with the terms of the client's contract with NIWA. This copyright extends to all forms of copying and any storage of material in any kind of information retrieval system.

Whilst NIWA has used all reasonable endeavours to ensure that the information contained in this document is accurate, NIWA does not give any express or implied warranty as to the completeness of the information contained herein, or that it will be suitable for any purpose(s) other than those specifically contemplated during the Project or agreed by NIWA and the Client.

	Formatting checked by:	
	Approved for release by:	

This report should be referenced:

Finucci, B., Stephenson, F., Petersen, G., Francis, M.P. & Pinkerton M.H. (2020). Exploring the drivers of spatial distributions of basking shark (*Cetorhinus maximus*) in the South Pacific. *NIWA Client Report prepared for Department of Conservation (POP2020-03)*: 53 p.

Contents

Executive summary	5
1 Introduction	7
2 Methods.....	9
2.1 Study area	9
2.2 Species records	10
2.3 Environmental and biotic predictor variables	10
2.4 Habitat Suitability Modelling (HSM)	12
3 Results	17
3.1 Basking shark records	17
3.2 Model performance	17
3.3 Variable selection and contribution	17
3.4 Predicted basking shark distributions.....	20
4 Discussion	22
4.1 Drivers of basking shark distribution	23
4.2 Basking shark habitat suitability in New Zealand	24
4.3 Future directions.....	25
5 Acknowledgements	26
6 References.....	27
Appendix A Environmental and biotic variables.....	36
Appendix B Spatial coverage of environmental variables	41
Appendix C Partial dependence plots.....	50

*The default Table of Contents shows two levels based on Headings 1 and 2 and the Numbered Headings 1 and 2. You can change the number of levels to be displayed by selecting **Change TOC Format** on the **NIWA CLIENT REPORT** tab.*

Tables

Table 1: Spatial environmental and biotic predictor variables included in the final models, collated for species distribution models from Stephenson et al. (2020). Further details for each environmental variable are available in Stephenson et al. (2020) and details on the biotic variables are available in Pinkerton et al.

	(2020). All other environmental and biotic predictor variables are found in Appendix A.	11
Table 3:	Mean cross-validated estimates of model performance for the bootstrapped boosted regression tree (BRT) and random forest (RF) models.	17
Figures		
Figure 1:	Map of the study region (New Zealand Exclusive Economic Zone (EEZ), black dashed line), bathymetry and feature names used throughout the text modified from Stephenson et al., 2020, and the location of basking shark records used in this study (black dots).	9
Figure 2:	The 95% kernel density estimate (KDE) probability grid (home-range). Areas outside the KDE probability grid are covered by crossed black lines.	13
Figure 3:	Pearson's correlation coefficients among the final environmental and biotic variables.	14
Figure 4:	Partial dependence plots of the mean boosted regression tree (BRT) and random forest (RF) models for the nine variables, showing the influence of each predictor variable on the response. Variables are ordered by influence as indicated in top left hand of plots. Shaded area represents 95% confidence interval.	19
Figure 5:	The predicted habitat suitability index (HSI) of basking shark in the New Zealand Exclusive Economic Zone (EEZ) modelled using the bootstrapped ensemble models. Areas outside 95% kernel density estimate (KDE) probability grid indicating lower confidence that can be placed in the predicted habitat suitability are covered by crossed black lines.	20
Figure 6:	The predicted habitat suitability index (HSI) of basking shark in the New Zealand Exclusive Economic Zone (EEZ) modelled using the bootstrapped ensemble models for A) West Coast South Island; B) East Coast South Island; C) south of South Island including Puysegur and Stewart Island; D) Chatham Islands; and E) Auckland Islands. Areas outside 95% kernel density estimate (KDE) probability grid indicating lower confidence that can be placed in the predicted habitat suitability are covered by crossed black lines. Note that the Chatham Islands (D) is outside the KDE probability grid estimate.	21
Figure 7:	Standard deviation of the predicted habitat suitability index (HSI) of basking shark in the New Zealand Exclusive Economic Zone (EEZ) modelled using the bootstrapped ensemble models. Areas outside 95% kernel density estimate (KDE) probability grid indicating lower confidence that can be placed in the predicted probability occurrence are covered by crossed black lines.	22

Executive summary

Historically, basking sharks have been widely reported throughout New Zealand waters. While previously observed in large numbers, only a few individuals are now reported annually, primarily as fisheries bycatch. Basking shark observations are known to be highly variable across years, and their distribution and occurrence in the Northern Hemisphere have been shown to be influenced by environmental predictors such as thermal fronts, chlorophyll *a* (chl-*a*) concentration, and the abundance of prey (zooplankton). It is unclear if the recent disappearance of basking sharks in New Zealand is a result of changes in encounter rates with different fisheries operational methods changes in regional distribution of sharks, or a true decline in basking shark abundance.

Correlative models that predict the occurrence of species in relation to environmental variables (termed species distribution models or habitat suitability models, HSMs) have become an important part of resource management and conservation biology. Such models are capable of filling in knowledge gaps on spatial and temporal distributions and predict areas of suitable habitat for widely distributed species. Here, basking shark habitat suitability (HSI) was predicted by combining functionally relevant, high-resolution environmental and biotic (prey species) data (1 km² grid resolution) and basking shark occurrence data (n = 369) that has been opportunistically recorded across New Zealand's Exclusive Economic Zone (EEZ).

The relationship between environment variables, biotic variables and basking shark records was explored using ensemble predictions (Ensemble HSM) from Boosted Regression Tree (BRT) and Random Forest (RF) models. BRT and RF models were bootstrapped 200 times and an ensemble model was produced by taking weighted averages of the predictions from each model type. BRT and RF models performed well for predicting basking shark occurrence (AUC and TSS > 0.7).

Nine variables were retained for the model, eight environmental predictors (*Bathy*, *BPI broad*, *Chl-a*, *MLD*, *Turbidity*, *POCFlux*, *Slope*, and *SST*) and one biotic predictor (*Copepoda*). The relative importance of each predictor and their influence on basking shark HSI were consistent across BRT and RF models. Vertical flux (*POCFlux*, 26.0%), slope (*Slope*, 14.1%), and turbidity (*Turbidity*, 10.6%) were the three most important variables in predicting basking shark HSI. Bathymetry (*Bathy*, 9.7%) and broadscale bathymetric position index (*BPI broad*, 9.6%) were also moderately important variables. High HSI was predicted in gently sloping and less complex seafloor topographies with low turbidity and at two depths - very close to shore and at depths between 200 and 550 m. There was a weak relationships between HSI and copepod densities, with low HSI occurring with low levels of copepod densities, a peak in HSI at moderate copepod densities (10-20 counts per 5 nautical miles), and a plateau in HSI values at the highest levels of copepod densities (>25 counts per 5 nautical miles). HSI was lowest at moderate levels of chl-*a* concentration (*Chl-a*) (0.5-1.0 mg m⁻³) and highest at high chl-*a* concentration (>1.2 mg m⁻³).

Areas of high habitat suitability for basking shark in New Zealand waters occurred along the continental slope, particularly along the 250 m contour along the North and South Islands; Mernoo Bank, Pukaki Rise, Puysegur, and around New Zealand's offshore islands (Chatham Islands, Stewart Island, Bounty Islands, and Auckland Islands). Areas of high uncertainty (SD > 0.2) included most offshore waters north of 40°S, the deeper depths (>500 m) of the Hokitika Canyon, northern Chatham Rise, coastal waters off East Coast South Island (Canterbury Bight), Foveaux Strait (between the South Island and Stewart Island) and Puysegur. High uncertainty beyond the core area was reported along deep sea features north of New Zealand, including the Kermadec Ridge and Trench, Colville Ridge, Norfolk Ridge, and Lord Howe Rise.

DRAFT

The outputs produced here will be useful for assessing threats and conservation needs (e.g. spatially explicit risk assessment), and providing guidance for future research efforts (e.g. areas of interest for sampling). This study has provided the first insight into habitat suitability for basking sharks, not only in the South Pacific, but in the Southern Hemisphere, using a novel approach by incorporating both environmental and biotic predictors into habitat models. However, caution should be considered given the relatively small sample size of occurrence data, lack of absence data, and potential spatial bias in sampling effort (use of fisheries dependent data).

1 Introduction

The basking shark (*Cetorhinus maximus*) is a planktivorous coastal-pelagic species widely distributed in the temperate and tropical waters of the Atlantic and Pacific Oceans, and fringes of the Indian Ocean (southern Australia, Indonesia, South Africa) (Rigby et al., 2019). It is the second largest fish in the world after the whale shark (*Rhincodon typus*), reaching an estimated maximum size of at least 10 m total length (Weigmann, 2016). Basking sharks are known for their slow surface swimming behaviour but are also capable of vertical migrations to depths of 1,264 m (Gore et al., 2008). The species also engages in long distance migrations and has been recorded crossing the eastern to western Atlantic Ocean (Braun et al., 2018; Dewar et al., 2018; Johnston et al., 2019). Habitat use and movement patterns in the Pacific Ocean, and more specifically around New Zealand, however, are virtually unknown. Despite its large size, basking sharks remain elusive and data-poor in the South Pacific.

Historically, basking sharks have been widely reported throughout New Zealand at latitudes between 39°S and 51°S. Most records are from south of Cook Strait, including a number of records from the brackish waters of Lake Ellesmere (Te Waihora) (Francis and Duffy, 2002). Individuals have been most commonly reported in nearshore habitat on the east and west coast of the South Island, and in waters around the Snares and Auckland Islands during the spring and summer months (Francis, 2017). Off Banks Peninsula, aerial surveys for Hector's dolphins (*Cephalorhynchus hectori*) conducted by the Department of Conservation reported large groups of over 100 individuals in the early 1990s (Francis and Duffy, 2002). Such large sightings have not been reported since; aerial surveys have failed to see basking sharks in recent years and only a few individuals are now reported annually, primarily as fisheries bycatch (Francis and Duffy, 2002; Francis, 2017).

Basking sharks are susceptible to exploitation from fishing due to their naturally low population sizes, presumed slow growth rates, and low reproductive rates (Francis, 2017). The species has been subject to targeted fishing throughout their range, retained for their meat, skin, cartilage, liver oil, and high value fins (Rigby et al., 2019). While most targeted fisheries ceased in the 2000s, basking sharks are still taken as bycatch by a number of fishing gear types (e.g., trawl, trammel net, set net), and are threatened by interactions with recreational vessels and commercial shipping due to the species' habit of spending time at the surface (Austin et al., 2019; Rigby et al., 2019). Population recovery has been low or negligible several decades after the cessation of fishing (Fowler et al., 2005). In 2002, basking sharks were listed in Appendix II of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES, (CITES, 2002), and in 2005, were listed in Appendices I and II in the Convention of Migratory Species (CMS). In 2019, basking sharks were assessed as globally Endangered by the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (Rigby et al., 2019).

Basking sharks have been protected in New Zealand waters since 2010. Within New Zealand, the species has been assessed as nationally Vulnerable under both IUCN Red List Criteria and the New Zealand Threat Classification System (NZTCS) (Duffy et al., 2018; Finucci et al., 2019). There are no specific management measures in place for basking sharks, apart from mandatory reporting of captures and the return of captured individuals to the sea. In recent years, the species is occasionally taken as bycatch in trawl and set net fisheries, with trawl bycatch typically occurring near or beyond the edge of the continental shelf (Francis and Smith, 2010; Francis, 2017). Knowledge of the species is reliant on fisheries observers. There are very little fisheries independent data available and estimates of basking shark bycatch likely underestimate the total New Zealand catches because they do not account for captures in unobserved set net fisheries and inshore trawl fisheries (Francis, 2017). Patterns in unstandardised bycatch rates imply basking sharks were captured in relatively large

numbers in the late 1980s and early 1990s, with peak bycatch occurring between 1988 and 1991 (Francis, 2017). Following this period, observed bycatch rates declined dramatically. On the East Coast South Island, raw catch-per-unit-effort (CPUE) peaked in 1991 at 81.9 sharks per 1000 tows to no reported sharks from 2005–2016 (Francis, 2017). It is unclear if the recent decline in basking shark records in New Zealand is a result of a change to fishing methods that are less likely to encounter basking sharks, changes in regional availability of sharks, or a true decline in basking shark abundance (Francis, 2017).

Basking shark observations are known to be highly variable across years, with gaps in regional sightings of up to 20 years (Dewar et al., 2018). Basking shark distribution and occurrence appears to be strongly linked to zooplankton/prey abundance at smaller spatial scales, but the drivers of broad scale distribution patterns are largely unknown (Sims, 2008). In the Northern Hemisphere, their distribution appears to be influenced by environmental predictors such as sea surface temperature (SST), thermal fronts, chl-a concentration, and the abundance of zooplankton (Cotton, 2005; Austin et al., 2019). Without sufficient information on species' distribution, habitat use, and migratory patterns, it is difficult to determine the cause of variability in abundance.

Correlative models that predict the occurrence of species in relation to environmental variables (termed species distribution models or habitat suitability models) have become an important part of resource management and conservation biology. Such models are capable of filling knowledge gaps on spatial and temporal distributions and predicting areas of suitable habitat for widely distributed species (Elith et al., 2006; Weber et al., 2017). By relating species' sightings to environmental predictor variables, the abundance or probability of taxa presence can be estimated along with a characterisation of the environmental drivers of species distributions. These models are becoming increasingly popular for use on marine species spanning large geographic and bathymetric ranges and have been employed on number of cetaceans (Stephenson et al., 2020), seabirds (Cleasby et al., 2020), and cartilaginous fishes, including basking sharks in the Northeast Atlantic (Austin et al., 2019).

Here, we predict basking shark habitat suitability by combining functionally relevant, high-resolution environmental and biotic (prey species) data (1 km² grid resolution) with available basking shark occurrence data opportunistically recorded across New Zealand's Exclusive Economic Zone (EEZ). The distribution of prey, included here as zooplankton densities, often-overlooked, and at times, a key predictor of species' distributions (Dormann et al., 2018). Understanding biotic interactions and their influence in driving species' distributions is important for predicting into unsampled space because the trophic interactions that are at the core of species habitat use may be better captured (e.g. more accurate predictions due to climate change) (Araújo and Luoto, 2007). Identifying factors driving basking shark distribution across the New Zealand marine region is important for better understanding species' regional ecology and direct and inform future research and spatially-focused conservation efforts of this protected species.

2 Methods

2.1 Study area

The study area extends over 4.2 million km² of the South Pacific Ocean within the New Zealand Exclusive Economic Zone (EEZ, $\approx 25 - 57^{\circ}\text{S}$; $162^{\circ}\text{E} - 172^{\circ}\text{W}$; Figure 1). New Zealand waters contain highly productive zones of mixing between higher salinity, nutrient poor, warm, northern waters, and lower salinity, nutrient rich, cold, southern water, resulting in areas of high biological diversity which are suitable for a range of shark species (Bradford-Grieve et al., 2006; Leathwick et al., 2006; Stephenson et al., 2018; Stephenson et al., 2020).

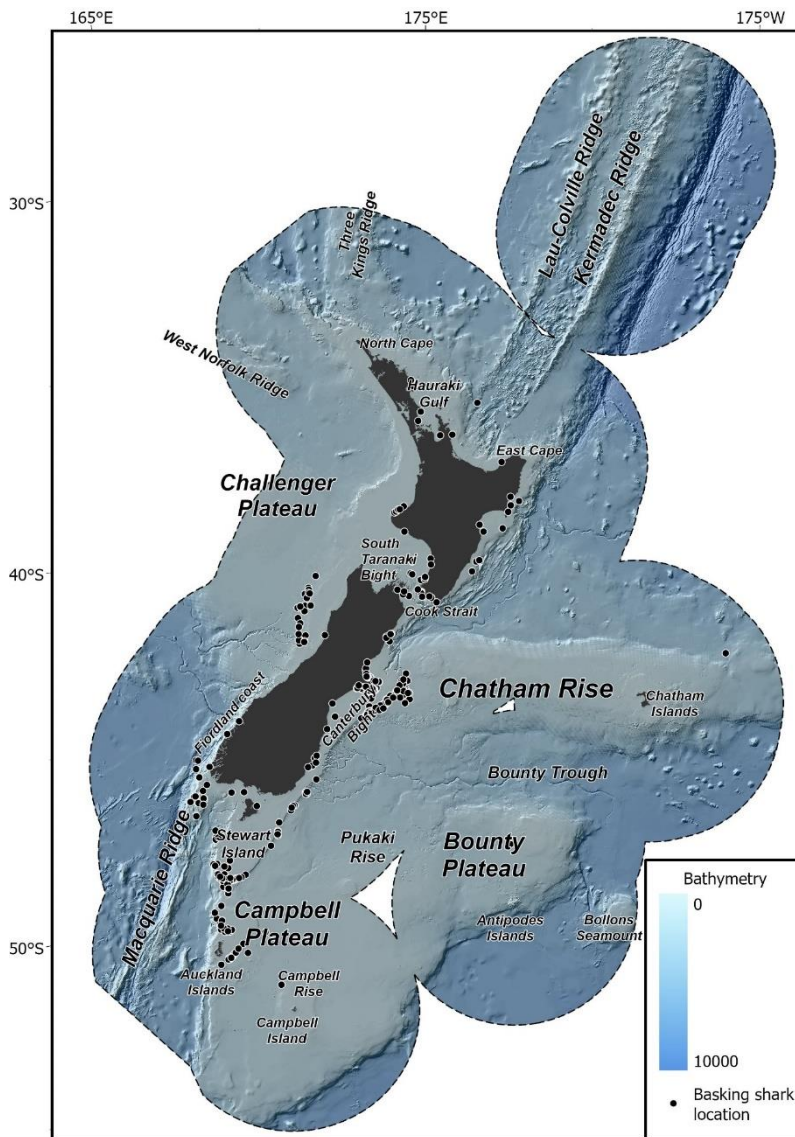


Figure 1: Map of the study region. (New Zealand Exclusive Economic Zone (EEZ), black dashed line), bathymetry and feature names used throughout the text modified from Stephenson et al., 2020, and the location of basking shark records used in this study (black dots).

2.2 Species records

Habitat Suitability Models (HSMs) were used to analyse and spatially predict the distribution of basking shark habitat suitability (measured as habitat suitability index – HSI). Basking shark records ($n = 401$) were collated from various sources (Francis and Duffy, 2002; Francis, 2017; C.A.J. Duffy, unpublished data) and included records from commercial fisheries and observations pooled from public sightings, media reports, museum records, scientific surveys, and beach cast specimens. Records included information on date, number of individuals, geographic co-ordinates and source (where available) and were collected between 1889 and 2020. The data were groomed in previous work (Francis and Duffy, 2002; Francis and Smith, 2010; Francis, 2017) to only keep records that were confirmed or probable basking shark observations that were within the New Zealand EEZ. The most recent records reported by fisheries observers were confirmed with photoidentification. Because of difficulties in correcting for differences in sampling methods, all catch records were converted into presence records (Elith et al., 2011; Stephenson et al., 2018). To minimize the effect of spatial bias in the occurrence data, species records were aggregated spatially to a 1 km grid resolution (Aiello-Lammens et al., 2015; Stephenson et al., 2020). Strandings and reports without an approximate date reference (month) were removed. The final dataset included presence records of basking sharks at 369 unique sampling locations.

2.3 Environmental and biotic predictor variables

To characterise variability in the New Zealand marine environment, a comprehensive dataset of spatial environmental variables were collated at a 1 km grid resolution, with each spanning the breadth of the New Zealand EEZ (Table 1 and Appendix A Table 1, further details are available in Stephenson et al. (2020)). In addition to environmental variables, spatial estimates of various zooplankton densities (Pinkerton et al., 2020) (inferred prey) were used as a biological predictor in the models (Appendix A, Table 1). Estimates of zooplankton densities did not cover the entire New Zealand EEZ (Appendix B, Figure 9). Areas lacking this information will simply represent the modelled relationship between basking shark records and the environmental variables. A preliminary examination of currently available zooplankton density estimates reveals these are likely to cover core areas of basking shark distribution. Of the available environmental and biotic variables, a subset was selected to be used in the SDMs (Table 1) based on model tuning described in section 2.4.2. Although most of the chosen environmental variables were static (e.g., bathymetry, Bathy), several variables were dynamic in time, representing mean monthly statistics (e.g., chlorophyll-a concentration, Chl-a, “temporal resolution” column in Table 1).

Prior to fitting of the habitat suitability models, values for each environmental and biotic variable were extracted for locations of basking shark records by overlaying the records onto each of the environmental and biotic variable layers using the “raster” package in R (Hijmans and van Etten, 2012). For dynamic environmental variables (mean monthly environmental variables), recorded dates of basking shark records were used to extract respective values from the month the record was made.

Table 1: Spatial environmental and biotic predictor variables included in the final models, collated for species distribution models from Stephenson et al. (2020). Further details for each environmental variable are available in Stephenson et al. (2020) and details on the biotic variables are available in Pinkerton et al. (2020). All other environmental and biotic predictor variables are found in Appendix A.

Abbreviation	Full name	Temporal resolution	Description	Units
<i>Bathy</i>	Bathymetry	Static	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)	m
<i>BPI_broad</i>	Bathymetric position index_broad	Static	Terrain metrics were calculated using an inner annulus of 12 km and a radius of 62 km using the NIWA bathymetry layer in the Benthic Terrain Modeler in ArcGIS 10.3.1.1 (Wright et al., 2012). Bathymetric Position Index (BPI) is a measure of where a referenced location is relative to the locations surrounding it.	m
<i>Chl-a</i>	Chlorophyll-a concentration	Mean monthly	A proxy for the biomass of phytoplankton present in the surface ocean (to ~30 m). Blended from a coastal Chl-a estimate (quasi-analytic algorithm (QAA), local aph*(555)) and the default open-ocean chl-a value from MODIS-Aqua (v2018.0) (Pinkerton, 2016)	mg m ⁻³
<i>MLD</i>	Mixed layer depth	Mean monthly	The depth that separates the homogenized mixed water above from the denser stratified water below. Based on GLBu0.08 hindcast results using a potential density difference of 0.030 kg m ⁻³ from the surface. Models used are: (1) hycom: from day 265 (2008) to present; (2) fnmoc: from day 169 (2005) to present; (3) soda: from day 249 (1997) to end of 2004; (4) tops: from day 001 (2005) to 225 (2010) (Pinkerton, 2016)	m
<i>POCFlux</i>	Downward vertical flux of particulate organic matter at the seabed	Mean monthly	Net primary production in the surface mixed layer estimated as the VGPM model (Behrenfeld and Falkowski, 1997); this table). Export fraction and flux attenuation factor with depth estimated by refitting sediment trap and thorium-based measurements to environmental data (VGPM, SST) as (Lutz et al., 2002; Pinkerton, 2016) and using data from (Cael et al., 2018).	mgC m ⁻² d ⁻¹

<i>Turbidity</i>	Particulate backscatter at 555 nm (previously used to generate 'turbidity')	Mean monthly	Optical particulate backscatter at 555 nm estimated using blended coastal and ocean products. Coastal: QAA v5 product bbp555 from MODIS-Aqua data. Ocean: <i>bbp_555_giop</i> ocean product (Werdell, 2019) . Result calculated as long-term (2002–2017) average.	m ⁻¹
<i>Slope</i>	Slope	Static	Bathymetric slope was calculated from water depth and is the degree change from one depth value to the next.	Degree
<i>SST</i>	Sea surface temperature	Mean monthly	Blended from OI-SST (Reynolds et al., 2002) ocean product and MODIS-Aqua SST coastal product. Long-term (2002–2017) average values at 250 m resolution.	°C
<i>Copepoda</i>	Copepoda	Static	Copepods, including calanoid, other cyclopoid, and harpacticoid copepods across at least 50 species. Most abundant identified species include <i>Calanus simillimus</i> (29%) and <i>Ctenocalanus citer</i> (27%) (Pinkerton et al., 2020).	Counts per 5 nautical mile Continuous Plankton Recorder (CPR) segment

2.4 Habitat Suitability Modelling (HSM)

The relationship between environment variables, biotic variables and basking shark records was explored using ensemble predictions (Ensemble HSM) from Boosted Regression Tree (BRT) and Random Forest (RF) models. This approach limits dependence on a single model type or structural assumption and enables a more robust characterization of the predicted spatial variation and uncertainties (Robert et al., 2016).

To estimate basking shark distributions, BRT and RF models require locations of both presences (occurrence records) and absences. Here, true absences (i.e. sample locations where no basking sharks were recorded) were not easily available for each sampling method (i.e. trawl tows, observer records, scientific surveys) and were not available for opportunistic records such as public sightings, media report, or museum records. Therefore, presence only modelling approaches using pseudo absences (i.e. locations where basking sharks were not recorded within our study area) was necessary.

2.4.1 Pseudo-absence selection

A two-dimensional kernel density estimate (KDE) was produced using all basking shark locations (presence data) (Figure 2) using a cell size of 1km². Within the KDE, the 95% percentage volume contour (minimum area in which 95% of the KDE value is located) was selected (referred to in some cases as the home-range, (Calenge, 2006)), and used to create a probability grid from which pseudo-absences were sampled according to the probability of grid weights (that is, where KDE values were high, the chance of selecting an absence was high) (Georgian et al., 2019). Pseudo-absences were generated through random selection of points from within the probability grid except within a 1 km-grid radius of the presence localities. By selecting pseudo-absences in this manner, the pseudo-absences were subject to the same sampling bias as the presence data. This method has been shown to significantly increase the accuracy of BRT and RF models (Elith et al., 2010; Cerasoli et al., 2017;

Georgian, 2019; Georgian et al., 2019). Following recommended best practice, the number of pseudo-absences selected by month were equivalent to the number of monthly presences (Barbet-Massin et al., 2012).

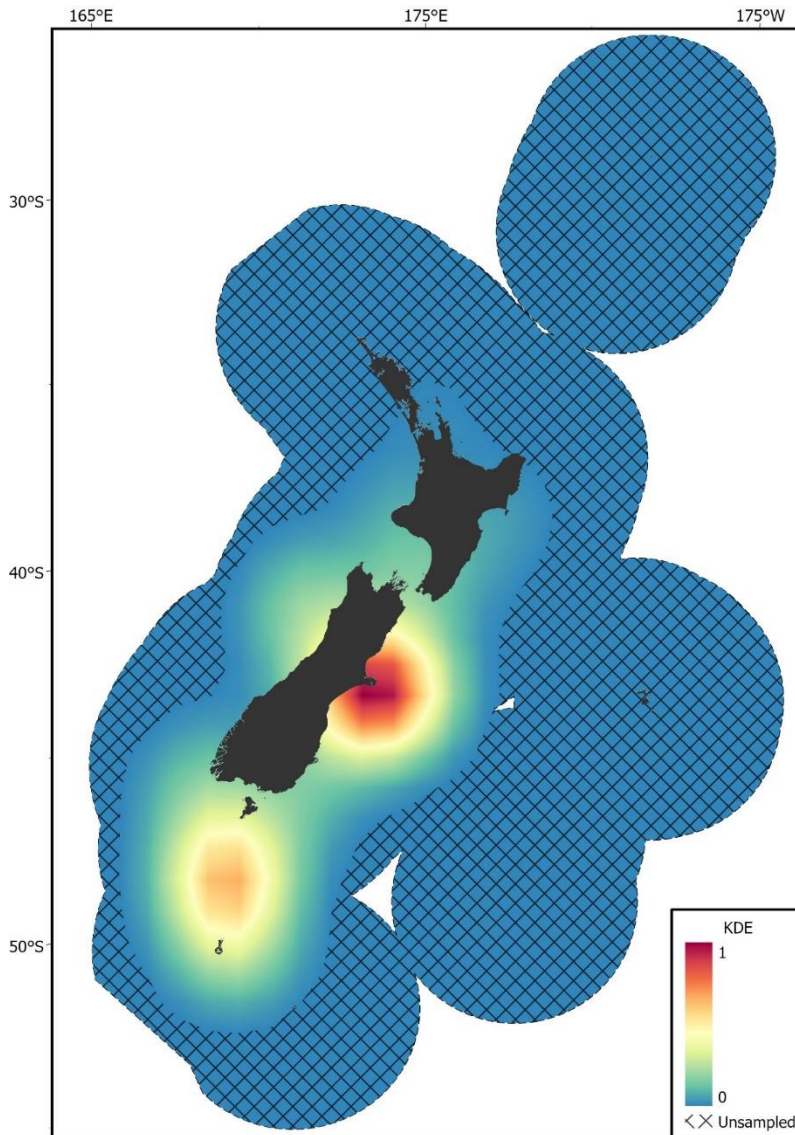


Figure 2: The 95% kernel density estimate (KDE) probability grid (home-range). Areas outside the KDE probability grid are covered by crossed black lines.

2.4.2 Predictor variable selection

In most cases, the inclusion of many variables (e.g. > 20 variables) in tree-based machine learning models (i.e., BRT and RF) is avoided because they only provide minimal improvement in predictive accuracy, and complicate interpretation of model outcomes (Leathwick et al., 2006). As the interpretation of drivers of distribution of basking shark was a key requirement, a reduction in the number of predictor variables was undertaken in order to produce a parsimonious model. A BRT model was initially fitted using all available environmental variables which was then 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. The model is then 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 the deviance explained decreased by > 1% between removal of predictor variables. Bottom temperature was initially found to be a moderately important predictor in the model but was later removed based on the observation that basking sharks are often captured in mid-water trawls and thus, bottom temperature may not be representative of the species' true habitat preference. Despite having a relatively small influence on the model, Chl-a was retained as this predictor was found to be an important predictor of basking shark distribution elsewhere (Austin et al., 2019).

The final variables retained for modelling were *Bathymetry*, *BPI broad*, *Chl-a*, mixed layer depth (*MLD*), *Turbidity*, *POCFlux*, *Slope*, sea surface temperature (*SST*), and Copepoda (Table 1). Several environmental variables showed some co-linearity (Figure 3) however, all levels of co-linearity were considered acceptable for tree-based machine learning methods (Pearson correlation < 0.75, (Elith et al., 2010; Dormann et al., 2013). The 'final' environmental variables selected through this approach were also used in RF models (Appendix B Figure 1-Figure 9).

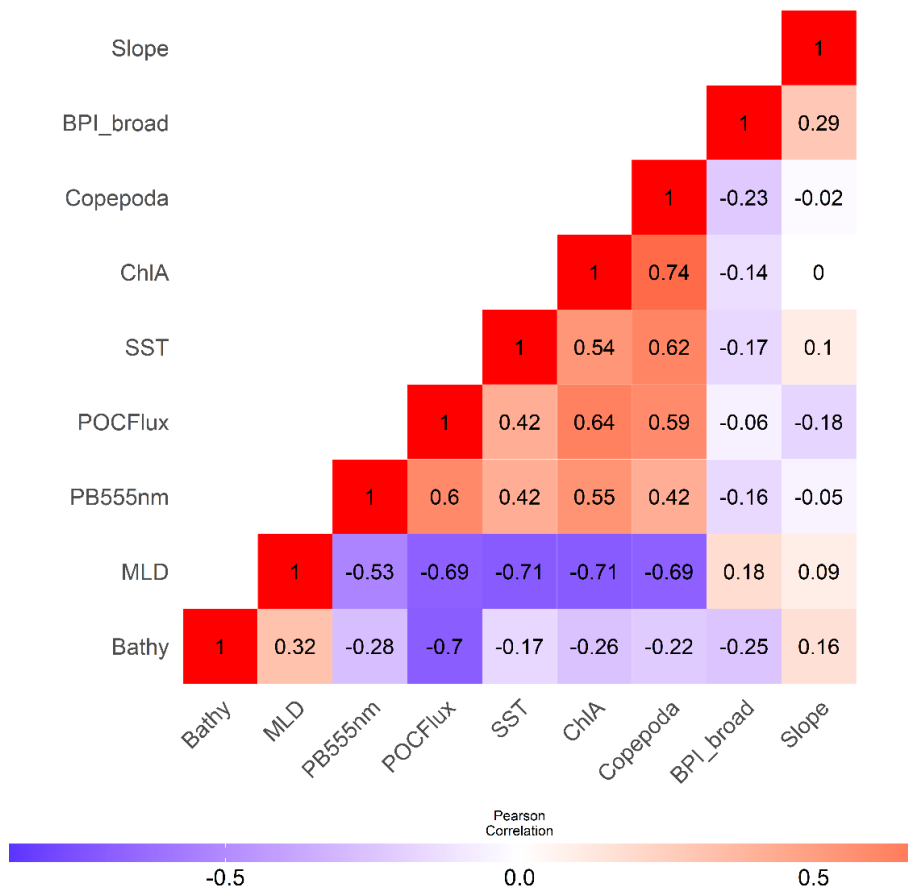


Figure 3: Pearson's correlation coefficients among the final environmental and biotic variables.

2.4.3 Boosted Regression Tree 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 et al., 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 et al., 2017). BRT models were fitted with a Bernoulli error distribution, a tree complexity of 2, a learning rate of 0.01 (with parameters selected so as to fit trees for each bootstrapped model), a bag fraction of 0.7 and random 10-fold cross evaluation following recommendations from Leathwick et al. (2006) and Elith et al. (2008). The BRT method has been widely used in ecological applications and has performed well in previous studies of fish and cetacean distributions in New Zealand (Leathwick et al., 2006; Compton et al., 2013; Stephenson et al., 2020)

2.4.4 Random Forest models

RF models (Breiman, 2001) fit an ensemble of regression (abundance data) or classification tree (presence/absence data) models describing the relationship between the distribution of an individual species and some set of environmental variables (Ellis et al., 2012). Following environmental and biotic predictor variable selection using the BRT model, the RF model was tuned using the train function in the R package 'caret' (Kuhn, 2020). This function selects optimal values for the complexity parameters *mtry* (the number of variables used in each tree node), *maxnodes* (the maximum number of terminal nodes in each trees), and *ntree* (the number of trees to grow). RF models have previously been applied to demersal fish in the New Zealand EEZ (Stephenson et al., 2018).

2.4.5 Bootstrapping the models

BRT and RF models were bootstrapped 200 times. A random 'training' sample consisting of the total presence records was drawn with replacement. A random sample of pseudo absence of equal number was drawn without replacement from the full set of available pseudo absences separated by month (Barbet-Massin et al., 2012) and the models were run using these presence-pseudo absence records. Presence records which were not randomly selected were combined with a random number of pseudo absences and were set aside for independent assessment of model performance (referred herein as 'evaluation' data). At each BRT and RF model iteration, geographic predictions were made using environmental predictor variables to a 1 km² grid. Given that BRT and RF models used pseudo absences, we refer to our outputs as 'habitat suitability' (rather than the commonly used probability of occurrence) because we did not have information on 'catchability' or 'sightability' of basking sharks from the different sampling methods nor did we have estimates of species prevalence (Anderson et al., 2016; Georgian et al., 2019). HSI and a spatially explicit measure of uncertainty (measured as the standard deviation of the mean, SD) were calculated for each grid cell using the 200 bootstrapped layers.

2.4.6 Model performance

BRT and RF model performance were evaluated using AUC (area under the Receiver Operating Characteristic curve) and TSS (True Skill Statistic). AUC is an effective measure of model performance and a threshold-independent measure of accuracy, while the TSS is a threshold-dependent measure of accuracy, but is not sensitive to prevalence (Allouche et al., 2006; Komac et al., 2016). AUC scores range from 0 – 1, with a score of 0.5 indicating model performance is equal to random chance, a score > 0.7 indicating adequate performance, and a score > 0.80 indicating excellent performance (Hosmer Jr et al., 2013). TSS, which takes into account Specificity and Sensitivity to provide an index ranging from -1 to +1, where +1 equals perfect agreement and -1 is no better than random, Allouche et al. (2006)). A TSS value > 0.6 is considered useful. (Allouche et al., 2006). Model fit metrics were calculated using both the 'training' dataset and the 'evaluation' dataset. The latter is considered a more robust

and conservative method of evaluating goodness-of-fit of a model than using the same data with which the model was trained (Friedman et al., 2001).

2.4.7 Ensemble models

We produced an ensemble model by taking weighted averages of the predictions from each model type, using methods adapted from Opper et al. (2012); Anderson et al. (2016); Rowden et al. (2017); Georgian et al. (2019); (Anderson et al., 2020). This adapted procedure derives a two-part weighting for each component of the ensemble model, taking equal contributions from the overall model performance (AUC value derived from the ‘evaluation’) and the uncertainty measure (SD) in each cell, as follows:

$$W1_{BRT} = \frac{MPS_{BRT}}{MPS_{BRT}+MPS_{RF}} \text{ and } W1_{RF} = \frac{MPS_{RF}}{MPS_{BRT}+MPS_{RF}}$$

$$W2_{BRT} = 1 - \frac{SD_{BRT}}{SD_{BRT}+SD_{RF}} \text{ and } W2_{RF} = 1 - \frac{SD_{RF}}{SD_{BRT}+SD_{RF}}$$

$$W_{BRT} = \frac{W1_{BRT}+W2_{BRT}}{2} \text{ and } W_{RF} = \frac{W1_{RF}+W2_{RF}}{2}$$

$$X_{ENS} = X_{BRT} * W_{BRT} + X_{RF} * W_{RF}$$

$$SD_{ENS} = SD_{BRT} * W_{BRT} + SD_{RF} * W_{RF}$$

where MPS_{BRT} and MPS_{RF} are the model performance statistics; X_{BRT} and X_{RF} are the model predictions; SD_{BRT} and SD_{RF} are the bootstrap SDs; and X_{ENS} and SD_{ENS} are the weighted ensemble predictions and weighted SDs, respectively, from which maps of predicted species distribution and model uncertainty were produced. All spatial outputs from this work are provided at a 1km grid resolution and using the Albers Equal Area projection centered at 175°E and 40°S (EPSG:9191), a standard format now accepted by the Department of Conservation (DOC) and Fisheries New Zealand (FNZ) (Wood et al., (in prep)).

Two measures of spatially explicit uncertainty were produced: an estimate of our spatial coverage of species occurrences (the home-range (95% KDE)) and the standard deviation of the predicted basking shark distribution (i.e. model uncertainty). The calculated home-range (spatial coverage of species occurrences) was assumed to be indicative of basking shark distributions, and thus, is presumed to have more certain predictions of the species’ distribution. Where predictions were projected outside the home-range (i.e. where there are few or no sightings), it is assumed that the relationship between the environment and species’ records may be less robust and thus predictions outside the home-range contain some degree of uncertainty (e.g., similarly to the methods used in (Stephenson et al., 2020)). Standard deviation (SD) of the mean predicted habitat suitability were estimated through the bootstrapping methods outlined in section 2.4.5 and are provided as uncertainty estimates of basking shark distribution.

Ensemble model performance was assessed using AUC and TSS by comparing ensemble model predictions to all basking shark presence records and an equal number of randomly selected pseudo absence data. To ensure that the random selection of pseudo absence data did not provide misleading model performance metrics, this procedure was iterated 50 times and mean AUC and TSS score calculated for the ensemble model (Barbet-Massin et al., 2012).

Partial dependence plots were made for the BRT and RF models to evaluate the effect of each predictor on species' distribution by plotting the effect of the predictor on the response (basking shark presence) after accounting for the average effects of all other model predictors (Elith et al., 2008). Ensemble partial dependence plots were created with an average of the BRT and RF partial dependence plots.

3 Results

3.1 Basking shark records

Most basking shark records (72%, $n = 265$) occurred in the spring and summer months (September to February). Since 2000, most records (84%, $n = 103$) have been from fishing events, with one aerial record and 19 opportunistic sightings. In the past decade, all but two of the 45 basking shark records were from fishing interactions.

3.2 Model performance

AUC and TSS scores using evaluation data were very similar between models, with the RF model performing slightly better than the BRT model (AUC: 0.92 and 0.89; TSS, 0.72 and 0.69 respectively, Table 2). Both indices indicated the models were useful in predicting basking shark occurrence (> 0.7). Measures of BRT and RF model performance scores had low variability (measured by the standard deviation of the mean), suggesting the models were consistently performing across bootstrap samples. Model fits between training data and evaluation data were similar, with model fits for the evaluation data slightly lower than the training data (as would be expected). The similarity of these fits provides some indication that the training data were not overfit in the models.

Table 2: Mean cross-validated estimates of model performance for the bootstrapped boosted regression tree (BRT) and random forest (RF) models.

	Deviance explained (training data)	Deviance explained (evaluation data)	TSS (training data)	TSS (evaluation data)	AUC (training data)	AUC (evaluation data)
BRT model	0.60 ± 0.03	0.36 ± 0.10	0.92 ± 0.02	0.69 ± 0.05	0.95 ± 0.01	0.89 ± 0.03
RF model	0.75 ± 0.02	0.52 ± 0.07	0.88 ± 0.02	0.72 ± 0.04	0.98 ± 0.00	0.92 ± 0.02

3.3 Variable selection and contribution

Nine variables were retained for the model, eight environmental predictors (*Bathy*, *BPI broad*, *Chl-a*, *MLD*, *Turbidity*, *POCFlux*, *Slope*, and *SST*) and one biotic predictor (*Copepoda*). The relative importance of each predictor and their influence on basking shark habitat suitability were consistent across BRT and RF models (Appendix C Figure 10, Figure 11). Vertical flux (*POCFlux*, 26.0%), slope (*Slope*, 14.1%), and turbidity (*Turbidity*, 10.6%) were the three most important variables in predicting basking shark habitat suitability (Figure 4). Bathymetry (*Bathy*, 9.7%) and BPI broad (*BPI broad*, 9.6%) were also moderately important variables. There was a strong positive relationship of predicted basking shark

HSI with vertical flux, highest in areas where vertical flux was $20 \text{ mgC m}^{-2} \text{ d}^{-1}$ or greater than what would be expected for the given depth. High HSI was predicted in gently sloping and less complex seafloor topologies with low turbidity. Two depths strata had high HSI - nearshore depths and depths between 200 and 550 m. A less clear relationship was observed between HSI and sea surface temperature (*SST*) and mixed layer depth (*MLD*), with low HSI occurring between temperatures of 12.5°C and 15°C and in areas where the mixed layer depth was approximately 75 m. There was a weak relationships between HSI and copepod (*Copepoda*) densities, with low HSI occurring with low levels of copepod densities, a peak in HSI at moderate copepod densities (10-20 counts per 5 nautical miles), and a plateau in HSI values at the highest levels of copepod densities (>25 counts per 5 nautical miles). HSI was lowest at moderate levels of chl-a concentration (*Chl-a*) ($0.5\text{-}1.0 \text{ mg m}^{-3}$) and highest at high chl-a concentration ($>1.2 \text{ mg m}^{-3}$).

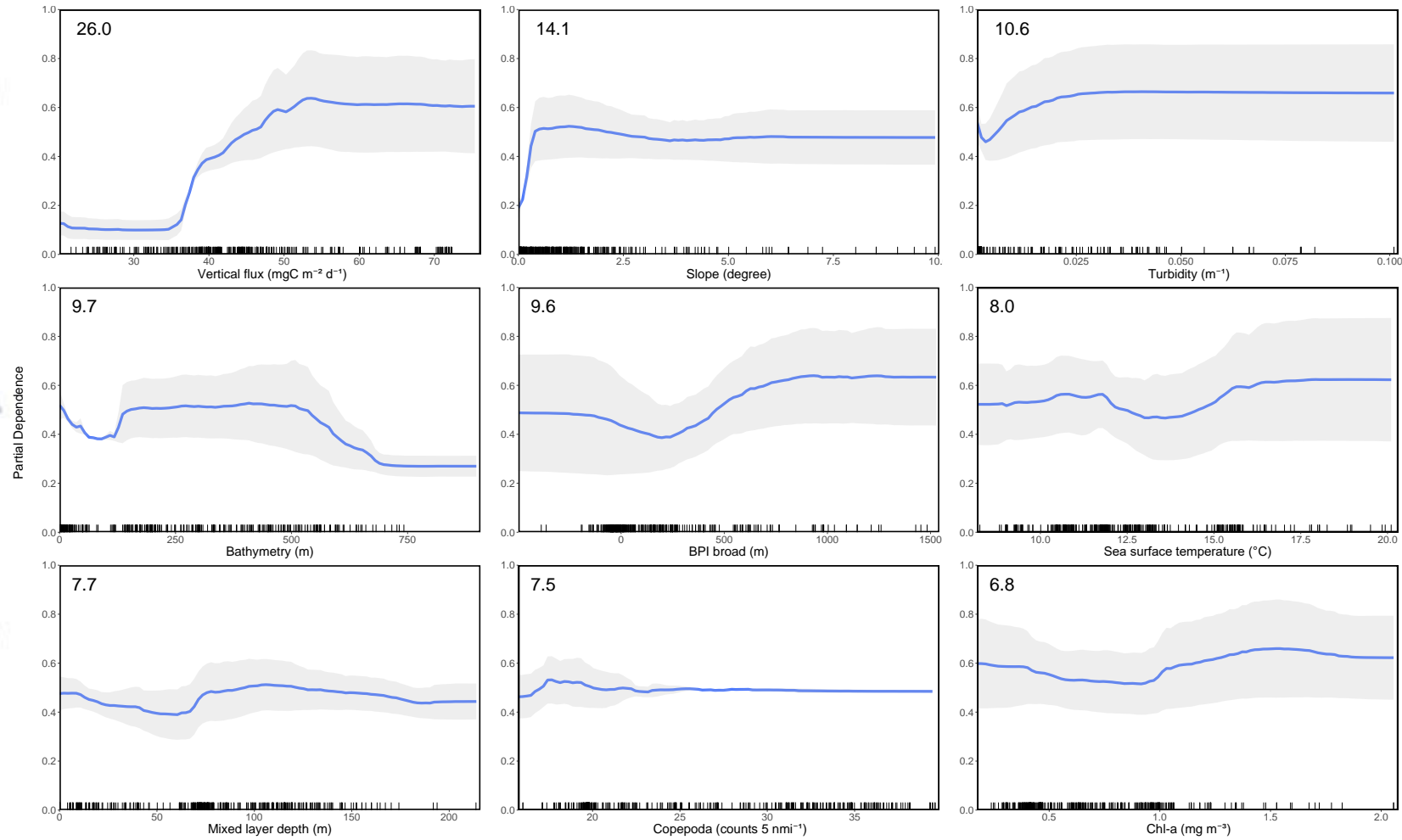


Figure 4: Partial dependence plots of the mean boosted regression tree (BRT) and random forest (RF) models for the nine variables, showing the influence of each predictor variable on the response. Variables are ordered by influence as indicated in top left hand of plots. Shaded area represents 95% confidence interval.

3.4 Predicted basking shark distributions

Areas of high habitat suitability for basking sharks in New Zealand waters occurred along the continental slope, particularly along the 250 m contour along the North and South Islands, Mernoo Bank, Pukaki Rise, Puysegur, and around New Zealand’s offshore islands (Chatham Islands, Stewart Island, Bounty Islands, and Auckland Islands) (Figure 5, Figure 6). Within the home-range, areas of high uncertainty ($SD > 0.2$) included most offshore waters north of $40^{\circ}S$, the deeper depths (>500 m) of the Hokitika Canyon, northern Chatham Rise, coastal waters off ECSI (Canterbury Bight), Foveaux Strait (between the South Island and Stewart Island) and Puysegur (Figure 7). The North Island and features further from the continental shelf, including Chatham Rise were outside of the estimated home-range. In addition, moderate - high uncertainty ($SD > 0.2$) beyond the home-range was reported along deep sea features north of New Zealand, including the Kermadec Ridge and Trench, Colville Ridge, Norfolk Ridge, and Lord Howe Rise (Figure 7).

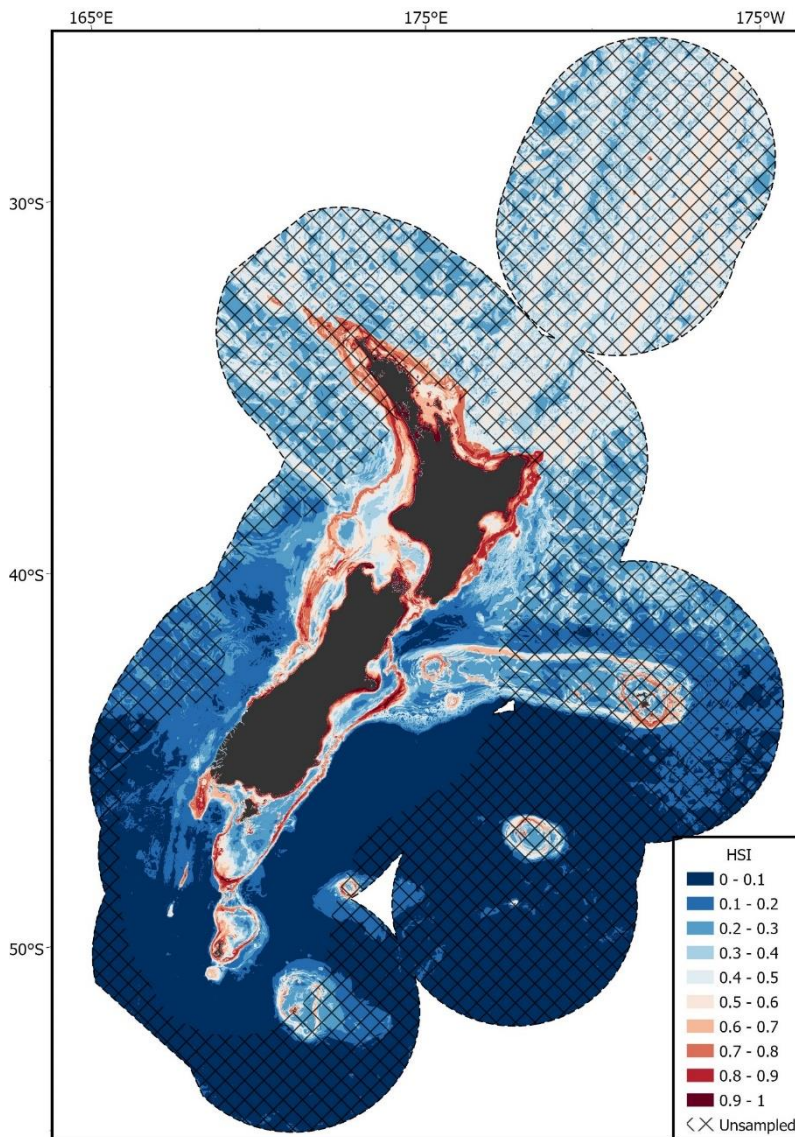


Figure 5: The predicted habitat suitability index (HSI) of basking shark in the New Zealand Exclusive Economic Zone (EEZ) modelled using the bootstrapped ensemble models. Areas outside 95% kernel density estimate (KDE) probability grid indicating lower confidence that can be placed in the predicted habitat suitability are covered by crossed black lines.

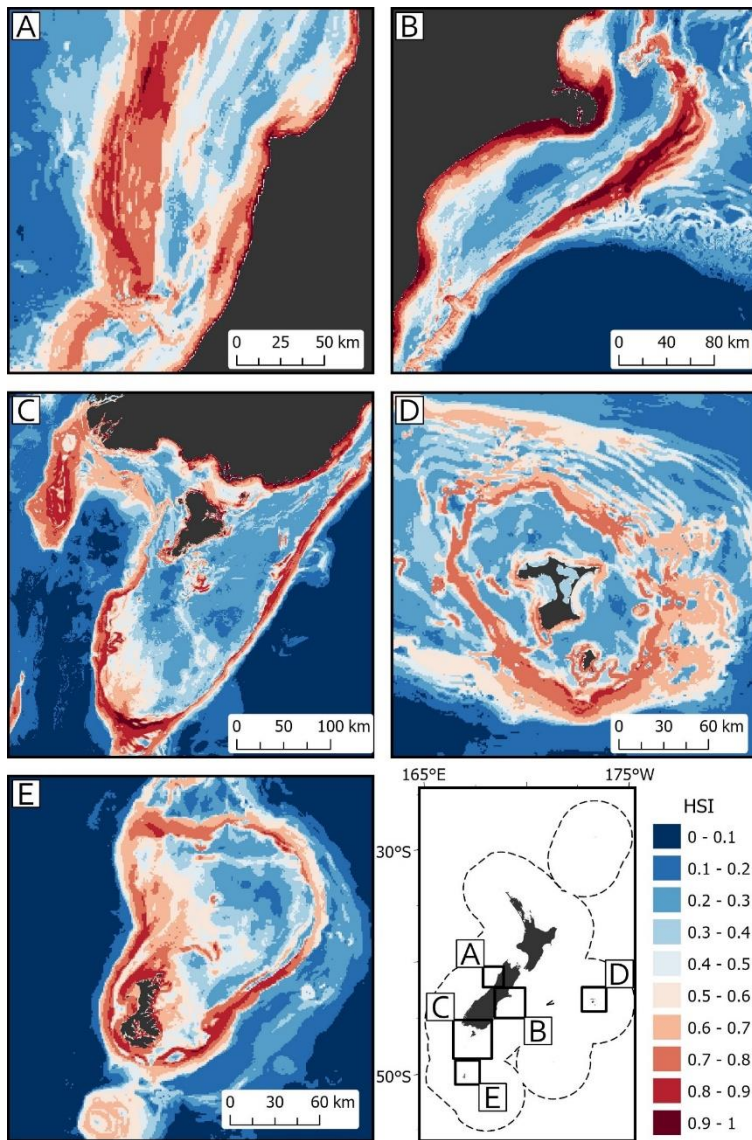


Figure 6: The predicted habitat suitability index (HSI) of basking shark in the New Zealand Exclusive Economic Zone (EEZ) modelled using the bootstrapped ensemble models for A) West Coast South Island; B) East Coast South Island; C) south of South Island including Puysegur and Stewart Island; D) Chatham Islands; and E) Auckland Islands. Areas outside 95% kernel density estimate (KDE) probability grid indicating lower confidence that can be placed in the predicted habitat suitability are covered by crossed black lines. Note that the Chatham Islands (D) is outside the KDE probability grid estimate.

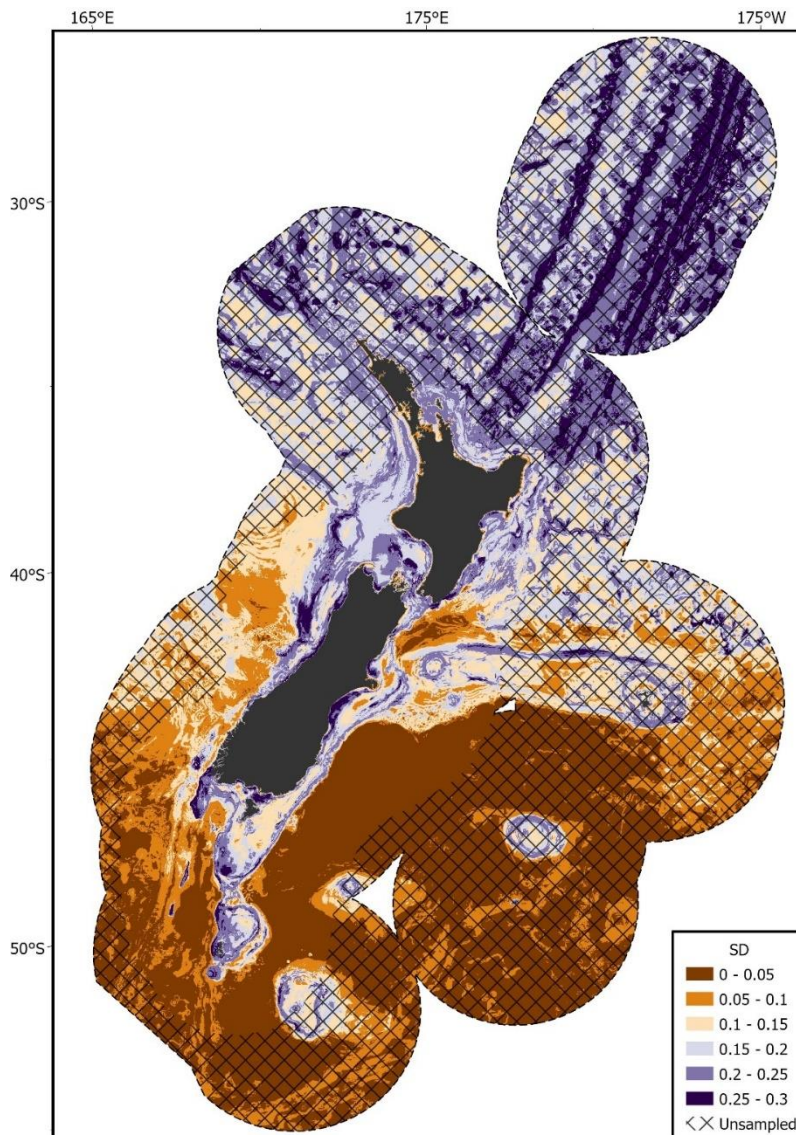


Figure 7: Standard deviation of the predicted habitat suitability index (HSI) of basking shark in the New Zealand Exclusive Economic Zone (EEZ) modelled using the bootstrapped ensemble models. Areas outside 95% kernel density estimate (KDE) probability grid indicating lower confidence that can be placed in the predicted probability occurrence are covered by crossed black lines.

4 Discussion

This study has provided the first insight into habitat suitability for basking sharks in the South Pacific and the Southern Hemisphere. Here, we have used a novel approach to assess habitat suitability by incorporating both temporally dynamic environmental ($n = 8$) and biotic ($n = 1$) predictors into the ensemble models. The BRT and RF models had good predictive power (AUC and TSS > 0.7) and both models performed similarly with low variability in the model fit metrics. The outputs produced here will be useful for assessing risk to fisheries and conservation needs (e.g. spatially explicit risk assessment), as well as providing guidance for future research efforts (e.g. areas of interest for sampling). However, caution should be considered given the relatively small sample size of species presences and lack of true absence data.

4.1 Drivers of basking shark distribution

Basking shark habitat suitability was largely influenced by variables representing ocean processes. Areas with high levels of vertical flux of particulate organic matter at the seabed (net primary production in the surface mixed layer) were most influential – which is likely indicative of high levels of primary production and prey density at the seafloor. In the Northeast Atlantic, basking sharks are often observed in shallow, highly productive coastal waters during spring and summer months where they feed on zooplankton blooms (Sims, 2008). Movement patterns within New Zealand waters are unknown but given the complexity of the regional marine environment and little population differentiation across global regions (Lieber et al., 2020), it is plausible basking sharks also partake in large migrations in New Zealand over prolonged periods of time. The inclusion of dynamic (mean monthly) environmental variables here may indicate seasonal patterns of basking shark distribution, with both inshore and offshore regions highlighted as areas of high habitat suitability. This is particularly evident in the bimodal effect of the bathymetry predictor, where basking shark habitat suitability was observed to be highest in very shallow depths (<100 m), and again at depths between 200 and 500 m.

While bathymetry (and slope) were also found to be important predictors, their effect may be partially influenced by fisheries availability (see below). Basking sharks have been shown to dive as deep as 1264 m and have been documented at depths of 600–1100 m (Francis and Duffy, 2002; Gore et al., 2008; Doherty et al., 2017). The species has also been shown to follow distinct water masses at depth, remaining at depths of 250 m or more for months without coming to the surface (Braun et al., 2018; Dewar et al., 2018). Basking sharks are known for complex diel vertical movements, which are thought to be influenced by shifts in prey availability and oceanography (Sims et al., 2005; Dewar et al., 2018). In well-stratified deep waters, basking sharks exhibit normal diel vertical movements (shallow depths at night, deeper depths during daylight), while sharks occupying inshore, inner-shelf areas near thermal fronts conducted reverse diel vertical movements (shallow depths during the day, deeper depths at night) (Sims et al., 2005). This may explain, at least in part, why more contemporary records occur in fisheries operating during daylight hours (when sharks are occurring at their preferred deeper depth range).

Thermal preference had relatively minimal influence on basking sharks. Basking sharks appear to have a broad thermal range and are therefore relatively unrestricted by temperature (Sims et al., 2003). They can cross tropical regions by submerging into deeper, colder water (Skomal et al., 2009) and one individual was encountered in tropical waters off Indonesia (Fahmi and White, 2015). While sea temperatures may have minimal effect on basking sharks, processes associated with sea surface temperatures that are expected to shift with climate change might. By 2100, climate change projections predict sea surface temperature will increase by 2.5°C, which in turn is predicted to lead to declines in surface mixed layer depth (by 15%), primary production (4.5%) and particle flux (12%); with the largest changes in macronutrients predicted in eastern Chatham Rise and southern Sub-Antarctic waters (Law et al., 2018). Such changes in the marine environment may not only reduce food availability for basking sharks, but may also alter their distribution. Basking shark movement patterns have been linked to shifts in prey availability and oceanography (Sims et al., 2005; Gore et al., 2008; Dewar et al., 2018). One tagged individual was shown to remain in area with putative upwelling and high abundance of phytoplankton in Western Atlantic regions for up to a month at a time (Gore et al., 2008). In the Northeast Atlantic, a northward shift in basking shark distribution in response to long-term zooplankton declines was found to correspond with declines in basking shark catch in Irish fisheries from 1948 to 1975 (Sims and Reid, 2002).

The biotic predictive layers included here were found to have lower influence on habitat suitability compared to some of the environmental predictors. Prey availability is highly patchy and temporally

variable; thus, it is possible a static variable reflecting prey abundance was unable to accurately represent the spatial distribution of prey. However, the inclusion of biotic predictors in the model is important in understanding species' relationship with the marine environment in unobserved space and has been identified as a potential link in understanding effects in climate change. Although prey preference for New Zealand sharks is unknown, there appears to be some relationship to copepod abundance, as seen in the North Atlantic (Sims and Merrett, 1997). Chl-a concentration, often used as an index of phytoplankton abundance (primary production) and strongly linked to primary consumers such as copepods, was predicted here to positively influence basking shark HSI. In recent decades, dramatic shifts in chl-a concentration have been reported the South Pacific and the Southern Oceans; significant declines were observed in spring and summer months in the South Pacific from 1979–2000 and significant increases linked to extreme summer marine heatwaves in the Southern Ocean between 2002 and 2018 (Gregg and Conkright, 2002; Montie et al., 2020). Similar models used in this project could be explored to predict basking shark distribution response to future climate change forecasting.

4.2 Basking shark habitat suitability in New Zealand

Areas of high basking shark habitat suitability included the east and west coasts of the South Island, Puysegur, and southern edge of Campbell Plateau. Some areas of Chatham Rise, specifically around Mernoo Bank and off the southern slope of Pitt Island (Chatham Islands), were also identified as areas of high habitat suitability. Much of Chatham Rise, however, was outside the home-range and thus holds a degree of uncertainty. Chatham Rise is a known hotspot for chondrichthyan diversity in New Zealand waters (Finucci et al. in prep), but interestingly, basking sharks have very rarely been reported from here. Star Keys, a group of five rocky islets approximately 12 kilometres east of Pitt Island, is a documented hotspot for white sharks (*Carcharodon carcharias*) (Duffy et al., 2012) and their presence may perhaps deter basking sharks from remaining in the region. Alternatively, Chatham Rise, as well as Puysegur, have relatively low densities of copepods (see Appendix B Figure 9), and may not be optimal feeding grounds for basking sharks.

Given the long temporal span (121 years) of the data, model predictions may be more representative of past, and not current, suitable habitat for basking sharks in New Zealand waters. Some predicted inshore habitat suitability was likely influenced by past inshore sightings, particularly around the North Island. Public sighting of basking sharks have not been reported around northern New Zealand since the early 2000s, despite regular sightings of other large marine megafauna in the region (e.g. (Duffy, 2002; Duffy and Tindale, 2018) and targeted offshore surveys (Zaeschmar et al., 2020). In addition, some inshore regions are unlikely to be attractive to basking sharks as they overlap with the distribution of globally known predators, such as white sharks in Kaipara harbour, orcas (*Orcinus orca*) (widespread around New Zealand), and sperm whales (*Physeter macrocephalus*) off Kaikoura (Francis, 1996; Sagnol et al., 2015; Stephenson et al., 2020).

There were a number of areas where the spatially explicit uncertainty (measured as the SD) was relatively high, indicating the relationship between basking sharks and the environment are more uncertain. These areas, such as the Cook Strait, northern Chatham Rise, and Foveaux Strait, are where few basking shark sightings were available. Uncertainties regarding the most northern predictions of habitat (north of 40°S) may, in part, be explained by a lack of information on copepod density north of 40°S (Pinkerton et al., 2020). Differences in habitat suitability among the sexes or size classes, a common observation among shark species, were not examined at this time due to the relatively small sample size of basking sharks across the region.

The estimate of spatial coverage of species occurrences (top 95% of the KDE of basking shark occurrences) provides a representation of the likely geographic (and in turn environmental) space

occupied by basking sharks within New Zealand waters. Predicted distribution outside of this ‘home-range’ area should be treated with caution as the prediction will not be underpinned by occurrence records and thus represents prediction into unsampled space. In this study, environmental threshold reflects the distribution of presences only – and thus retain any spatial biases associated with these datasets. In particular, the spatial distribution of presences will be related to the distribution of fishing effort and human population centres (for opportunistic sightings) and may not be accurate representation of hotspots. However, using the top 95% of the KDE of basking shark occurrences as an estimate of home-range provides a conservative estimate of the species space use that is useful in determining when modelled predictions are occurring outside of sampled environmental space. This provides a meaningful threshold with which to classify broad areas as ‘uncertain’.

4.3 Future directions

The lack of basking shark records in recent years highlights the need to better understand the disappearance or lack of detection of the species from New Zealand waters. However, without dedicated surveys and research efforts, and the paucity of fisheries-independent data, current records are reliant on interactions with fisheries (Francis, 2017). Our knowledge of New Zealand basking shark distribution occurs in areas of relatively high historic and current trawl fishing effort (Baird et al., 2018). Most basking shark interactions occur during the spring-summer months, corresponding to when fishing vessels target commercially important species, such as spawned cohorts of arrow squid (*Nototodarus sloanii*) (Hurst et al., 2012). As a protected species, it is mandatory to report basking shark interactions, but observer coverage is relatively low in some sectors, particularly within inshore fisheries and it is suspected that these interactions are underestimated (Francis, 2017). Understanding habitat use will assist in assessing risk to fishing activities and could be incorporated into management frameworks such as spatially explicit risk assessment (Large et al., 2019).

Identifying areas of high habitat suitability will assist in decision making processes for future research efforts. Previous research has identified the need to tag free-swimming basking sharks to better understand species movement, habitat use, and interactions with fisheries (Francis, 2017). To do so, this will require the ability to find individuals at the surface in an accessible location. The total South Pacific basking shark population size is unlikely to be high; in the Northeast Atlantic, basking shark numbers likely do not exceed 10,000 individuals (Lieber et al., 2020). This may make species’ detection more difficult in the vast marine space of New Zealand’s EEZ. Aerial surveys have successfully detected New Zealand basking sharks previously, and such surveying has been useful for estimating regional population sizes (Francis and Duffy, 2002; Westgate et al., 2014). However, if basking sharks are not feeding, they may travel in subsurface habitat and may therefore go undetected (Hawkes et al., 2020). By identifying areas of high habitat suitability, research efforts can be directed to specific areas of interest. For example, the Auckland Islands has been identified with high habitat suitability for basking sharks. This area is also known to be a hotspot for southern right whales (*Eubalaena australis*) during the Austral winter months (Rayment et al., 2015). Southern right whales follow the Subtropical Front (STF), a continuous feature within the Southern Tropical Convergence at latitudes 39° – 42°S, characterized by elevated primary productivity (Mackay et al., 2020). Southern Ocean oceanographic fronts have been identified as important foraging areas for a range of marine predators (Bost et al., 2009) and may also be important for basking sharks. More at-sea distribution data is required to understand habitat use, threat overlap, and population status throughout the New Zealand and South Pacific region.

5 Acknowledgements

This project was carried out under the Department of Conservation project POP2020-03. Thanks to Jade Maggs (NIWA) for the data extraction, to Clinton Duffy (DOC), Ben Sharp, and Marco Milardi (FNZ) for their feedback on the methods, and thanks to Tom Brough for reviewing the draft report.

6 References

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., and Anderson, R. P. (2015). spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* **38**: 541-545.

Allouche, O., Tsoar, A. and Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of applied ecology* **43**(6): 1223-1232.

Anderson, A., Stephenson, F., and Behrens, E. (2020). Updated habitat suitability modelling for protected corals in New Zealand waters, NIWA report prepared for Department of Conservation (DOC) NIWA CLIENT REPORT No: 2020174WN

Anderson, O. F., Guinotte, J. M., Rowden, A. A., Tracey, D. M., Mackay, K. A. and 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.

Araújo, M. B., and Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography* **16**: 743-753.

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

Baird, S. J., and Wood, B. A. (2018). Extent of bottom contact by New Zealand commercial trawl fishing for deepwater Tier 1 and Tier 2 target fishstocks, 1989–90 to 2015–16. *New Zealand Aquatic Environment and Biodiversity Report No. 193*: 102.

Barbet-Massin, M., Jiguet, F., Albert, C. H. and Thuiller, W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* **3**(2): 327-338.

Behrenfeld, M. J., and Falkowski, P. G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and oceanography* **42**: 1-20.

Bost, C. A., Cotté, C., Bailleul, F., Cherel, Y., Charrassin, J. B., Guinet, C., Ainley, D. G. a. and Weimerskirch, H. (2009). The importance of oceanographic fronts to marine birds and mammals of the southern oceans. *Journal of Marine Systems* **78**(3): 363-376.

Bradford-Grieve, J., Probert, K., Lewis, K., Sutton, P., Zeldis, J. and Orpin, A. (2006). New Zealand shelf region. *The sea* **14**.

Braun, C. D., Skomal, G. B. and Thorrold, S. R. (2018). Integrating Archival Tag Data and a High-Resolution Oceanographic Model to Estimate Basking Shark (*Cetorhinus maximus*) Movements in the Western Atlantic. *Frontiers in marine science* **5**.

Breiman, L. (2001). Random forests. *Machine learning* **45**(1): 5-32.

Cael, B., Bisson, K., and Follett, C. L. (2018). Can rates of ocean primary production and biological carbon export be related through their probability distributions? *Global biogeochemical cycles* **32**: 954-970.

Calenge, C. (2006). The package "adehabitat" for the R software: a tool for the analysis of space and habitat use by animals. *Ecological modelling* **197**: 516-519.

CARS2009 (2009, 10/04/2014). "CSIRO Atlas of Regional Seas." Retrieved 14/12/2017, 2017, from www.cmar.csiro.au/cars.

Cerasoli, F., Iannella, M., D'Alessandro, P. a. and Biondi, M. (2017). Comparing pseudo-absences generation techniques in Boosted Regression Trees models for conservation purposes: A case study on amphibians in a protected area. *PLoS ONE* **12**(1): p.e0187589.

CITES (2002). Consideration of proposals for amendment of Appendices I and II. Proposal: inclusion of Basking Shark (*Cetorhinus maximus*) on Appendix II of CITES. Prop. 12.36.

Cleasby, I. R., Owen, E., Wilson, L., Wakefield, E. D., O'Connell, P. and Bolton, M. (2020). Identifying important at-sea areas for seabirds using species distribution models and hotspot mapping. *Biological Conservation* **241**: p.108375.

Compton, T. J., Bowden, D. A., Roland Pitcher, C., Hewitt, J. E. and Ellis, N. (2013). Biophysical patterns in benthic assemblage composition across contrasting continental margins off New Zealand. *Journal of Biogeography* **40**(1): 75-89.

Cotton, P. A., Sims, D.W., Fanshawe, S., Chadwick, M. (2005). The effects of climate variability on zooplankton and basking shark (*Cetorhinus maximus*) relative abundance off southwest Britain.

Dewar, H., Wilson, S. G., Hyde, J. R., Snodgrass, O. E., Leising, A., Lam, C. H., Domokos, R., Wraith, J. A., Bograd, S. J., Van Sommeran, S. R. and Kohin, S. (2018). Basking Shark (*Cetorhinus maximus*) Movements in the Eastern North Pacific Determined Using Satellite Telemetry. *Frontiers in marine science* **5**.

Doherty, P. D., Baxter, J. M., Gell, F. R., Godley, B. J., Graham, R. T., Hall, G., Hall, J., Hawkes, L. A., Henderson, S. M., Johnson, L., Speedie, C. and Witt, M. J. (2017). Long-term satellite tracking reveals variable seasonal migration strategies of basking sharks in the north-east Atlantic. *Nature scientific reports* **7**:42837: 1-10.

Dormann, C. F., Bobrowski, M., Dehling, D. M., Harris, D. J., Hartig, F., Lischke, H., Moretti, M. D., Pagel, J., Pinkert, S., Schleuning, M., Schmidt, S. I., Sheppard, C. S., Steinbauer, M. J., Zeuss, D. and Kraan, C. (2018). Biotic interactions in species distribution modelling: 10 questions to guide interpretation and avoid false conclusions. *Global Ecology and Biogeography* **27**: 1004-1016.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B. and Leitão, P. J. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**(1): 27-46.

Duffy, C., Francis, M., Dunn, M., Finucci, B., Ford, R., Hitchmough, R. and Rolfe, J. (2018). Conservation status of New Zealand chondrichthyans (chimaeras, sharks and rays), 2016. New Zealand threat classification series: 13.

Duffy, C. A. and Tindale, S. C. (2018). First observation of the courtship behaviour of the giant devil ray *Mobula mobular* (Myliobatiformes: Mobulidae). *New Zealand Journal of Zoology* **45**(4): 387-394.

Duffy, C. A. J. (2002). Distribution, seasonality, lengths, and feeding behaviour of whale sharks (*Rhincodon typus*) observed in New Zealand waters. *New Zealand journal of marine and freshwater research* **36**: 565-570.

Duffy, C. A. J., Francis, M. P., Manning, M. and Bonfil, R. (2012). *Regional population connectivity, oceanic habitat, and return migration revealed by satellite tagging of white sharks, Carcharodon carcharias, at New Zealand aggregation sites*. Global perspectives on the biology and life history of the white shark. Domeier, M. L. Boca Raton, USA, CRC Press: 301-318.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J., Townsend, P., A, Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J., Williams, S., Wisz, M. S. and Zimmermann, N. E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**(2): 129-151.

Elith, J., Kearney, M. and Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution* **1**(4): 330-342.

Elith, J., Leathwick, J. R. and Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology* **77**(4): 802-813.

Elith, J., Leathwick, J. R. and Hastie, T. (2008). A working guide to boosted regression trees. *Journal of animal ecology* **77**: 802-813.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., and Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and distributions* **17**(1): 43-57.

Ellis, N., Smith, S. J. and Pitcher, C. R. (2012). Gradient forests: calculating importance gradients on physical predictors. *Ecology* **93**(1): 156-168.

Fahmi and White, W. T. (2015). First record of the basking shark *Cetorhinus maximus* (Lamniformes: Cetorhinidae) in Indonesia *Marine biodiversity records* **8**: e18.

Finucci, B., Duffy, C. A. J., Francis, M. P., Gibson, C. and Kyne, P. M. (2019). The extinction risk of New Zealand chondrichthyans. *Aquatic conservation: marine and freshwater ecosystems*.

Folk, R. L., Andrews, P. B. and Lewis, D. W. (1970). Detrital sedimentary rock classification and nomenclature for use in New Zealand. *New Zealand Journal of Geology and Geophysics* **13**(4): 937-968.

Fowler, S. L., Cavanagh, R. D., Camhi, M., Burgess, G. H., Cailliet, G. M., Fordham, S. V., Simpfendorfer, C. A. and Musick, J. A., Eds. (2005). *Sharks, rays and chimaeras: the status of chondrichthyan fishes*. Gland, Switzerland, IUCN.

Francis, M. P. (1996). *Observations on a pregnant white shark with a review of reproductive biology*. Great white sharks. The biology of *Carcharodon carcharias*. Klimley, A. P. and Ainley, D. G. San Diego, Academic Press: 157-172.

Francis, M. P. (2017). Review of commercial fishery interactions and population information for New Zealand basking shark. NIWA client report: 44.

Francis, M. P. and Duffy, C. (2002). Distribution, seasonal abundance and bycatch of basking sharks (*Cetorhinus maximus*) in New Zealand, with observations on their winter habitat. *Marine biology* **140**: 831-842.

Francis, M. P. and Smith, M. H. (2010). Basking shark (*Cetorhinus maximus*) bycatch in New Zealand fisheries, 1994–95 to 2007–08. New Zealand aquatic environment and biodiversity report: 57.

Friedman, J., Hastie, T. and Tibshirani, R. (2001). *The elements of statistical learning*, Springer series in statistics New York.

Frouin, R., McPherson, J., Ueyoshi, K., and Franz, B. A. (2012). A time series of photosynthetically available radiation at the ocean surface from SeaWiFS and MODIS data. *Remote Sensing of the Marine Environment II*: 852519.

Georgian, S. E., Anderson, O. F. and Rowden, A. A. (2019). Ensemble habitat suitability modeling of vulnerable marine ecosystem indicator taxa to inform deep-sea fisheries management in the South Pacific Ocean. *Fisheries Research* **211**: 256-274.

- Georgian, S. E., Anderson, O.F. and Rowden, A.A. (2019). Ensemble habitat suitability modeling of vulnerable marine ecosystem indicator taxa to inform deep-sea fisheries management in the South Pacific Ocean. *Fisheries research* **211**: 256-274.
- Gore, M. A., Rowat, D., Hall, J., Gell, F. R. and Ormond, R. F. (2008). Transatlantic migration and deep mid-ocean diving by basking shark. *Biology letters* **4**: 395–398.
- Gregg, W. W. a. and Conkright, M. E. (2002). Decadal changes in global ocean chlorophyll. *Geophysical Research Letters*, *29*(15), pp.20-1. **29**(15): 20-21-20-24.
- Hawkes, L. A., Exeter, O., Henderson, S. M., Kerry, C., Kukulya, A., Rudd, J., Whelan, S., Yoder, N. a. and Witt, M. J. (2020). Autonomous underwater videography and tracking of basking sharks. *Animal biotelemetry* **8**(1): 1-10.
- Hijmans, R. J., Phillips, S., Leathwick, J. and Elith, J. (2017). dismo: Species Distribution Modeling R package version 1.1-4. <https://CRAN.R-project.org/package=dismo>.
- Hijmans, R. J. and van Etten, J. (2012). raster: Geographic analysis and modeling with raster data. R package version 2.0-12. Available at: <http://CRAN.R-project.org/package=raster>.
- Hosmer Jr, D. W., Lemeshow, S. and Sturdivant, R. X. (2013). *Applied logistic regression*, John Wiley & Sons.
- Hurst, R. J., Ballara, S. L., MacGibbon, D. and and Triantafillos, L. (2012). Fishery characterisation and standardised CPUE analyses for arrow squid (*Nototodarus gouldi* and *N. sloanii*), 1989-90 to 2007-08, and potential management approaches for southern fisheries. New Zealand fisheries assessment report. **47**: 303.
- Johnston, E. M., Mayo, P. A., Mensink, P. J., Savetsky, E. and Houghton, J. D. R. (2019). Serendipitous re-sighting of a basking shark *Cetorhinus maximus* reveals inter-annual connectivity between American and European coastal hotspots. *Journal of fish biology* **95**(6): 1530-1534.
- Komac, B., Esteban, P., Trapero, L. and Caritg, R. (2016). Modelization of the current and future habitat suitability of *Rhododendron ferrugineum* using potential snow accumulation. *PLoS one* **11**(1): e0147324.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Team, R.C. and Benesty, M. (2020). Package ‘caret’. *The R Journal*.
- Large, K., Roberts, J., Francis, M. and Webber, D. N. (2019). Spatial assessment of fisheries risk for New Zealand sea lions at the Auckland Islands.
, New Zealand Aquatic Environment and Biodiversity Report. **No. 224**: 85.

D

Law, C. S., Rickard, G. J., Mikaloff-Fletcher, S. E., Pinkerton, M. H., Behrens, E., Chiswell, S. M. and Currie, K. (2018). Climate change projections for the surface ocean around New Zealand. *New Zealand journal of marine and freshwater research* **52**(3): 309-335.

R

Leathwick, J., Elith, J., Francis, M., Hastie, T. and 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.

A

Leathwick, J. R., Elith, J., Francis, M. P., Hastie, T. and 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.

F

Lieber, L., Hall, G., Hall, J., Berrow, S., Johnston, E., Gubili, C., Sarginson, J., Francis, M., Duffy, C., Wintner, S. P. and and Doherty, P. D. (2020). Spatio-temporal genetic tagging of a cosmopolitan planktivorous shark provides insight to gene flow, temporal variation and site-specific re-encounters. *Scientific reports* **10**(1): 1-17.

T

Lutz, M., Dunbar, R., and and Caldeira, K. (2002). Regional variability in the vertical flux of particulate organic carbon in the ocean interior. *Global biogeochemical cycles* **16**: 11-11.

Mackay, A. I., Bailleul, F., Carroll, E. L., Andrews-Goff, V., Baker, C. S., Bannister, J., Boren, L., Carlyon, K., Donnelly, D. M., Double, M. a. and Goldsworthy, S. D. (2020). Satellite derived offshore migratory movements of southern right whales (*Eubalaena australis*) from Australian and New Zealand wintering grounds. *PLoS ONE* **15**(5): p.e0231577.

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

Montie, S., Thomsen, M. S., Rack, W. a. and Broady, P. A. (2020). Extreme summer marine heatwaves increase chlorophyll a in the Southern Ocean. *Antarctic Science* **32**(6): 508-509.

Morel, A., Huot, Y., Gentili, B., Werdell, P. J., Hooker, S. B., and and Franz, B. A. (2007). Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach. *Remote Sensing of Environment* **111**: 69-88.

Oppel, S., Meirinho, A., Ramírez, I., Gardner, B., O'Connell, A. F., Miller, P. I. and Louzao, M. (2012). Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. *Biological Conservation* **156**: 94-104.

Pinkerton, M. (2016). Ocean colour satellite observations of phytoplankton in the New Zealand EEZ, 1997–2016. *Prepared for the Ministry for the Environment. Wellington: NIWA*.

Pinkerton, M. H., Décima, M., Kitchener, J. A., Takahashi, K. T., Robinson, K. V., Stewart, R. and Hosie, G. W. (2020). Zooplankton in the Southern Ocean from the continuous plankton recorder: Distributions and long-term change. *Deep Sea Research Part I: Oceanographic Research Papers* **162**.

Pinkerton, M. H., Décima, M., Kitchener, J. A., Takahashi, K. T., Robinson, K. V., Stewart, R. and Hosie, G. W. (2020). Zooplankton in the Southern Ocean from the continuous plankton recorder: Distributions and long-term change. *Deep Sea Research Part I: Oceanographic Research Papers*: 103303.

Pinkerton, M. H., Richardson, K. M., Boyd, P. W., Gall, M. P., Zeldis, J., Oliver, M. D., and and Murphy, R. J. (2005). Intercomparison of ocean colour band-ratio algorithms for chlorophyll concentration in the Subtropical Front east of New Zealand. *Remote Sensing of Environment* **97**: 382-402.

R Core Team (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria.

Rayment, W., Dawson, S., and and Webster, T. (2015). Breeding status affects fine-scale habitat selection of southern right whales on their wintering grounds. *Journal of biogeography* **42**(3): 463-474.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C. and Wang, W. (2002). An improved in situ and satellite SST analysis for climate. *Journal of climate* **15**: 1609-1625.

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

Ridgway, K., Dunn, J. and Wilkin, J. (2002). Ocean interpolation by four-dimensional weighted least squares—Application to the waters around Australasia. *Journal of atmospheric and oceanic technology* **19**(9): 1357-1375.

Rigby, C. L., Barreto, R., Carlson, J., Fernando, D., Fordham, S., Francis, M. P., Herman, K., Jabado, R. W., Liu, K. M., Marshall, A., Romanov, E. and Kyne, P. M. (2019). "*Cetorhinus maximus* (errata version published in 2020)." from <https://dx.doi.org/10.2305/IUCN.UK.2019-3.RLTS.T4292A166822294.en>.

Robert, K., Jones, D. O., Roberts, J. M. and Huvenne, V. A. (2016). Improving predictive mapping of deep-water habitats: Considering multiple model outputs and ensemble techniques. *Deep Sea Research Part I: Oceanographic Research Papers* **113**: 80-89.

Rowden, A. A., Anderson, O. F., Georgian, S. E., Bowden, D. A., Clark, M. R., Pallentin, A. and Miller, A. (2017). High-Resolution Habitat Suitability Models for the Conservation and Management of Vulnerable Marine Ecosystems on the Louisville Seamount Chain, South Pacific Ocean. *Frontiers in Marine Science* **4**: 335.

Sagnol, O., Richter, C., Reitsma, F. a. and Field, L. H. (2015). Estimating sperm whale (*Physeter macrocephalus*) daily abundance from a shore-based survey within the Kaikoura submarine canyon, New Zealand. *New Zealand journal of marine and freshwater research* **49**(1): 41-50.

Sims, D. W. (2008). Sieving a living: a review of the biology, ecology and conservation status of the plankton-feeding basking shark *Cetorhinus maximus*. *Advances in marine biology* **54**: 171-220.

Sims, D. W. and Merrett, D. A. (1997). Determination of zooplankton characteristics in the presence of surface feeding basking sharks *Cetorhinus maximus*. *Marine ecology progress series* **158**: 297-302.

Sims, D. W. and Reid, P. C. (2002). Congruent trends in long-term zooplankton decline in the north-east Atlantic and basking shark (*Cetorhinus maximus*) fishery catches off west Ireland. . *Fisheries oceanography* **11**: 59-63.

Sims, D. W., Southall, E. J., Richardson, A. J., Reid, P. C. and Metcalfe, J. D. (2003). Seasonal movements and behaviour of basking sharks from archival tagging: no evidence of winter hibernation. *Marine ecology progress series* **248**: 187-196.

Sims, D. W., Southall, E. J., Tarling, G. A. and Metcalfe, J. D. (2005). Habitat-specific normal and reverse diel vertical migration in the plankton-feeding basking shark. *Journal of animal ecology* **74**(4): 755-761.

Skomal, G. B., Zeeman, S. I., Chisholm, J. H., Summers, E. L., Walsh, H. J., McMahon, K. W. and Thorrold, S. R. (2009). Transequatorial migrations by basking sharks in the western Atlantic Ocean. *Current biology* **19**: 1019-1022.

Stephenson, F., Bulmer, R., Leathwick, J., Brough, T., Clark, D., Greenfield, B., Bowden, D., Tuck, I., Hewitt, J., Neill, K., Mackay, K., Pinkerton, M., Anderson, O., Gorman, R., Mills, S., Watson, S., Nelson, W., Rowden, A. and Lundquist, C. (2020). Development of a New Zealand Seafloor Community Classification (SCC). NIWA report prepared for Department of Conservation (DOC). Hamilton, National Institute of Water & Atmospheric Research.

Stephenson, F., Goetz, K., Sharp, B. R., Mouton, T. L., Beets, F. L., Roberts, J., MacDiarmid, A. B., Constantine, R., Lundquist, C. J. and Sarmiento Cabral, J. (2020). Modelling the spatial distribution of cetaceans in New Zealand waters. *Diversity and distributions* **26**(4): 495-516.

Stephenson, F., Leathwick, J. R., Francis, M. P. and Lundquist, C. J. (2020). A New Zealand demersal fish classification using Gradient Forest models. *New Zealand Journal of Marine and Freshwater Research* **54**(1): 60-85.

Stephenson, F., Leathwick, J. R., Geange, S. W., Bulmer, R. H., Hewitt, J. E., Anderson, O. F., Rowden, A. A. and 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.

D

Walters, R. A., Goring, D. G. and Bell, R. G. (2001). Ocean tides around New Zealand. *New Zealand Journal of Marine and Freshwater Research* **35**(3): 567-579.

R

Weber, M. M., Stevens, R. D., Diniz-Filho, J. A. F. and Grelle, C. E. V. (2017). Is there a correlation between abundance and environmental suitability derived from ecological niche modelling? A meta-analysis. *Ecography* **40**(7): 817-828.

A

Weigmann, S. (2016). Annotated checklist of the living sharks, batoids and chimaeras (Chondrichthyes) of the world, with a focus on biogeographical diversity. *Journal of fish biology*.

F

Werdell, P. J. (2019). Inherent Optical Properties (IOPs) Algorithm Theoretical Basis Document.

Westgate, A. J., Koopman, H. N., Siders, Z. A., Wong, S. N. P. a. and Ronconi, R. A. (2014). Population density and abundance of basking sharks *Cetorhinus maximus* in the lower Bay of Fundy, Canada. *Endangered species research* **23**: 177-185.

T

Wood, B., Finucci, B. and Black, J. ((in prep)). A standardised approach for creating spatial grids for New Zealand marine environment and species data.

Wright, D., Pendleton, M., Boulware, J., Walbridge, S., Gerlt, B., Eslinger, D., Sampson, D., and Huntley, E. (2012). ArcGIS Benthic Terrain Modeler (BTM), v. 3.0. *Environmental Systems Research Institute, NOAA Coastal Services Center, Massachusetts Office of Coastal Zone Management*.

Zaeschmar, J. R., Tezanos-Pinto, G., Dwyer, S. L., Peters, C. H., Berghan, J., Donnelly, D., Meissner, A. M., Visser, I. N., Weir, J. S., Judkins, A. G. a. and Brough, T. (2020). Occurrence, site fidelity, and associations of oceanic common bottlenose dolphins (*Tursiops truncatus*) off northeastern New Zealand. *Marine mammal science* **36**: 1180-1195.

Appendix A Environmental and biotic variables

Table 1: Spatial environmental and biotic predictor variables collated for species distribution models from Stephenson et al. (2020) and not included in the final model. Further details for each environmental variable are available in Stephenson et al. (2020) and details on the biotic variables are available in Pinkerton et al. (2020).

Abbreviation	Full name	Temporal resolution	Description	Units	Reference
<i>Beddist</i>	Benthic sediment disturbance	Static	One-year mean value of friction velocity derived from (1) hourly estimates of surface wave statistics (significant wave height, peak wave period) from outputs of the NZWAVE_NZLAM wave forecast, at 8-km resolution, (2) median grain size (d50), at 250 m resolution, (3) water depth, at 25-m resolution. Benthic sediment disturbance from wave action was assumed to be zero where depth \geq 200m.	ms^{-1}	
<i>BotNi</i>	Bottom nitrate	Static	Annual average water nitrate concentration at the seafloor (using NZ bathymetry layer) based on methods from (Reynolds et al., 2002). The oceanographic data used to generate these climatological maps were computed by objective analysis of all scientifically quality-controlled historical data from the Commonwealth Scientific and Industrial Research Organisation (CSIRO) Atlas of Regional Seas database (CARS2009, 2009).	umol l^{-1}	
<i>BotOxy</i>	Dissolved oxygen at depth	Static	Annual average water oxygen concentration at the seafloor (using NZ bathymetry layer) based on methods from (Reynolds et al., 2002). Oceanographic data from (CARS2009, 2009).	ml l^{-1}	
<i>BotPhos</i>	Bottom phosphate	Static	Annual average water phosphate concentration at the seafloor (using NZ bathymetry layer) based on methods from (Reynolds et al., 2002). Oceanographic data from (CARS2009, 2009).	umol l^{-1}	
<i>BotSal</i>	Salinity at depth	Static	Annual average water salinity concentration at the seafloor (using NZ bathymetry layer) based on methods from (Reynolds et al., 2002). Oceanographic data from (CARS2009, 2009).	psu	

<i>BotSil</i>	Bottom silicate	Static	Annual average water silicate concentration at the seafloor (using NZ bathymetry layer) based on methods from (Reynolds et al., 2002). Oceanographic data from (CARS2009, 2009).	$\mu\text{mol l}^{-1}$	
<i>BotTemp</i>	Temperature at depth	Static	Annual average water temperature at the seafloor (using NZ bathymetry layer) based on methods from (Ridgway et al., 2002). Oceanographic data from (CARS2009, 2009).	$^{\circ}\text{C km}^{-1}$	
<i>BPI_fine</i>	BPI_fine	Static	Terrain metrics were calculated using an inner annulus of 2 km and a radius of 12 km using the NIWA bathymetry layer in the Benthic Terrain Modeler in ArcGIS 10.3.1.1 (Wright et al. 2012). Bathymetric Position Index (BPI) is a measure of where a referenced location is relative to the locations surrounding it.	m	
<i>Chl-a.Grad</i>	Chlorophyll-a concentration spatial gradient	Mean monthly	Smoothed magnitude of the spatial gradient of annual mean Chl-a. Derived from Chl-a described above.	$\text{mg m}^{-3} \text{ km}^{-1}$	
<i>DET</i>	Detrital absorption	Mean monthly	Total detrital absorption coefficient at 443 nm, including due to coloured dissolved organic matter (CDOM) and particulate detrital absorption. Estimated using quasi-analytic algorithm (QAA) applied to MODIS-Aqua data, blended with <i>adg_443_giop</i> ocean product (Werdell, 2019).	m^{-1}	
<i>Ebed</i>	Seabed incident irradiance	Mean monthly	Broadband (400–700 nm) incident irradiance ($\text{E m}^{-2} \text{ d}^{-1}$) at the seabed, averaged over a whole year. Estimated by combining incident irradiance at the sea surface (Frouin et al., 2012) ; this table), diffuse downwelling irradiance attenuation (K_{PAR} ; this table) and bathymetric depth at monthly resolution. Derived from blended coastal (QAA) and open-ocean attenuation products.	$\text{E m}^{-2} \text{ d}^{-1}$	
<i>Kpar</i>	Diffuse downwelling attenuation	Mean monthly	vertical attenuation of diffuse, downwelling broadband irradiance (Photosynthetically Available Radiation, PAR, 400–700 nm). Merged coastal and open-ocean product based on MODIS-Aqua data. Coastal: estimated from inherent optical properties (QAA). Ocean:	m^{-1}	

			estimated from K490 using (Morel et al., 2007).		
PAR	Photo-synthetically active radiation	Mean monthly	Daily-integrated, broadband, incident irradiance at the sea-surface based on day length, solar elevation and measurements of cloud cover from ocean colour satellites (Frouin et al., 2012).	Einsteins m ⁻² d ⁻¹	
<i>SeasTDiff</i>	Annual amplitude of sea floor temperature	Static	Smoothed difference in seafloor temperature between the three warmest and coldest months. Providing a measure of temperature amplitude through the year.	°C km ⁻¹	
<i>Sed.class</i>	Sediment classification	Static	Classification of Mud, Sand and Gravel layers (this table) using the well-established (Folk et al., 1970) classification. Subtidal rocky reefs (this table) were incorporated. This classification provides a broad measure of hardness Mud – Rock.	NA; Mud; Muddy gravel; Muddy sandy gravel; sand; Gravelly mud; Gravelly sandy mud; Gravelly sand; Gravel; Rock	
<i>SstGrad</i>	Sea surface temperature gradient	Mean monthly	Smoothed magnitude of the spatial gradient of annual mean SST. This indicates locations in which frontal mixing of different water bodies is occurring (Leathwick et al., 2006). Derived from SST described above at two resolutions and merged.	°C km ⁻¹	
<i>SuspPM</i>	Suspended particulate matter	Mean monthly	Indicative of total suspended particulate matter concentration. Based on SeaWiFS ocean colour remote sensing data (Pinkerton et al., 2005); modified Case 2 atmospheric correction; modified Case 2 inherent optical property algorithm	Indicative of total suspended particulate matter concentration (g m ⁻³)	
<i>TC</i>	Tidal Current speed	Static	Maximum depth-averaged (NZ bathymetry) flows from tidal currents calculated from a tidal model for New Zealand waters (Walters et al., 2001). Tidal constituents (magnitude A and phase phi, represented as real and imaginary parts $X + iY = A \cdot \exp(i \cdot \phi)$) for sea surface height and currents (8 components) were taken from the EEZ tidal model, on an unstructured mesh at variable spatial resolution. The complex	ms ⁻¹	

			components were bilinearly interpolated to the output grid.		
TempRes	Temperature residuals	Static	Residuals from a GLM relating temperature to depth using natural splines – this highlights areas where average temperature is higher or lower than would be expected for any given depth	°C	
VGPM	Net primary production by the vertically-generalised production model	Mean monthly	Daily production of organic matter by the growth of phytoplankton in the surface mixed layer, net of phytoplankton respiration. Estimated at monthly resolution based on satellite observations of chl-a, PAR and SST, and model-derived estimates of mixed-layer depth, using the vertically-generalised production model (Behrenfeld and Falkowski, 1997).	mgC m ⁻² d ⁻¹	
Oithona	<i>Oithona similis</i>	Static	Cyclopoid copepods, dominated by <i>Oithona similis</i> (97%). The remaining 3% is unidentified (Pinkerton et al., 2020).	Counts per 5 nautical mile Continuous Plankton Recorder (CPR) segment	
Euphausiidae	Euphausiidae	Static	All adult and developmental stages of krill (generally not identified to species or genus). Most abundant identified species was <i>Thysanoessa macrura</i> (64%) (Pinkerton et al., 2020).	Counts per 5 nautical mile Continuous Plankton Recorder (CPR) segment	
Foraminifera	Foraminifera	Static	Unidentified (97.8%) Foraminifera specimens (Pinkerton et al., 2020).	Counts per 5 nautical mile Continuous Plankton Recorder (CPR) segment	
Fritillaria spp.	<i>Fritillaria</i> spp.	Static	Solitary, free-swimming larvacean, unidentified beyond genus (Pinkerton et al., 2020).	Counts per 5 nautical mile Continuous Plankton Recorder (CPR) segment	
Pteropods	Pteropods	Static	Pelagic gastropods, predominately <i>Limacina</i> spp. (98.9%) (Pinkerton et al., 2020).	Counts per 5 nautical mile Continuous Plankton Recorder (CPR) segment	

Zooplankton	Zooplankton	Static	Total abundance of all zooplankton types, including <i>Oithona similis</i> , Copepoda, Amphipoda, Chaetognatha, Euphausiidae, Foraminifera, <i>Fritillaria</i> spp., <i>Oikopleura</i> spp., Ostracoda, Pteropods, and "Other" (remaining identified organisms such as cephalopods and fish eggs comprising <1% of samples) (Pinkerton et al., 2020).	Counts per 5 nautical mile Continuous Plankton Recorder (CPR) segment	
-------------	-------------	--------	---	---	--

Appendix B Environmental predictor variables

Figure 1: Bathymetry (*Bathy*) within the New Zealand Exclusive Economic Zone (EEZ).

D
R
A
F
T

Figure 2: BPI broad (*BPI broad*) within the New Zealand Exclusive Economic Zone (EEZ).

Figure 3: Annual mean chlorophyll-a concentration (*Chl-a*) within the New Zealand Exclusive Economic Zone (EEZ).

D
R
A
F
T

Figure 4: Annual mean mixed layer depth (*MLD*) within the New Zealand Exclusive Economic Zone (EEZ).

Figure 5: Slope (*Slope*) within the New Zealand Exclusive Economic Zone (EEZ).

Figure 6: Annual mean sea surface temperature (SST) within the New Zealand Exclusive Economic Zone (EEZ).

Figure 7: Annual mean turbidity within the New Zealand Exclusive Economic Zone (EEZ).

Figure 8: Annual mean downward vertical flux of particulate (*POCflux*) within the New Zealand Exclusive Economic Zone (EEZ).

Figure 9: Modelled average copepod (subclass Copepoda) density, averaged for three times of day, six months (October to March) and years 1998–2018 from Pinkerton et al. (2020), reprojected for the New Zealand Exclusive Economic Zone (EEZ). Areas shown white either have no data, or no predictions were made, including because environmental conditions were outside the training data.

Appendix C Partial dependence plots

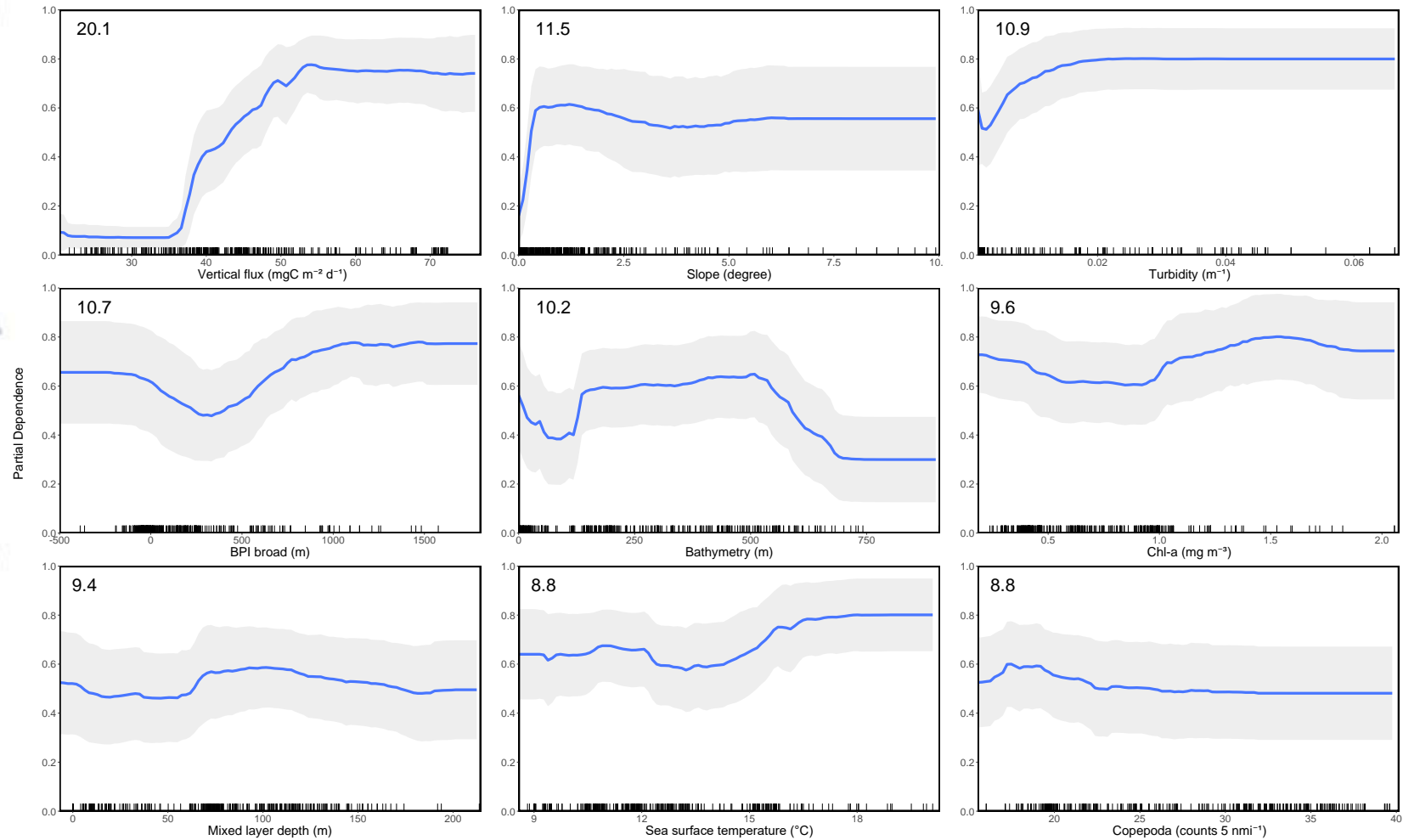


Figure 10: Partial dependence plots of the mean boosted regression tree (BRT) models for the nine variables, showing the influence of each predictor variable on the response. Variables are ordered by influence as indicated in top left hand of plots. Shaded area represents 95% confidence interval.

D
R
A
F
T

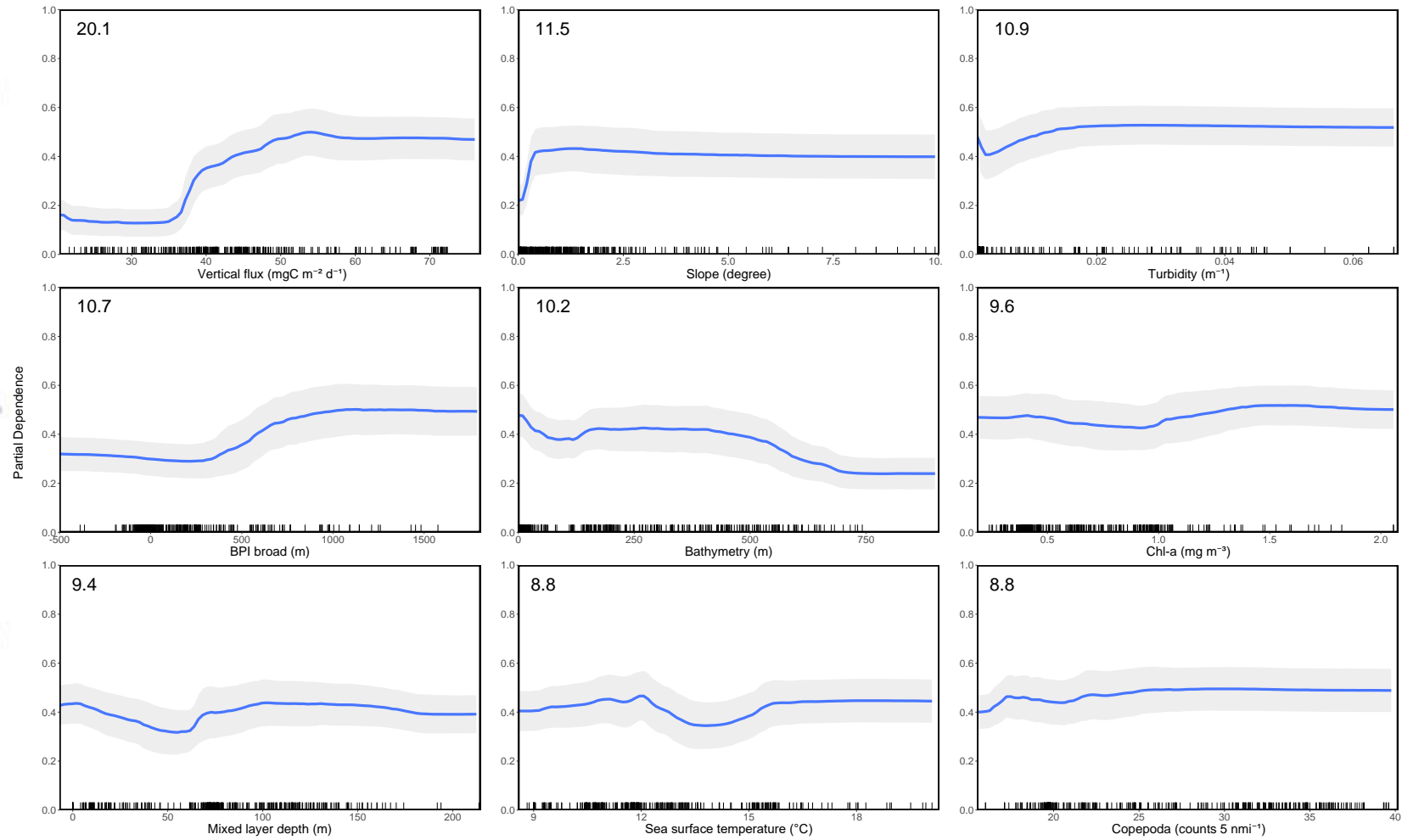


Figure 11: Partial dependence plots of the mean random forest (RF) models for the nine variables, showing the influence of each predictor variable on the response. Variables are ordered by influence as indicated in top left hand of plots. Shaded area represents 95% confidence interval.