Fisheries and Oceans Canada | Pêches et Océans Canada
Ecosystems and Oceans Science | Sciences des écosystèmes et des océans

Canadian Science Advisory Secretariat (CSAS)

Research Document 2020/004

Pacific Region

Development of a Species Distribution Modelling Framework and its Application to Twelve Species on Canada's Pacific Coast

Nephin[1], J., E.J. Gregr[2], C. St. Germain[1], C. Fields[1], J.L. Finney[1]

[1] Fisheries and Oceans Canada
Pacific Biological Station
3190 Hammond Bay Road
Nanaimo, BC V9T 6N7

[2] SciTech Environmental Consulting
2136 Napier Street
Vancouver, BC, V5L 2N9

March 2020

Canada

Foreword

This series documents the scientific basis for the evaluation of aquatic resources and ecosystems in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Published by:

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca

ISSN 1919-5044

Correct citation for this publication until published:

Nephin, J., Gregr, E.J., St. Germain, C., Fields, C., and Finney, J.L. 2020. Development of a Species Distribution Modelling Framework and its Application to Twelve Species on Canada's Pacific Coast. DFO Can. Sci. Advis. Sec. Res. Doc. 2020/004. xii + 107 p.

Aussi disponible en français :

Nephin, J., Gregr, E.J., St. Germain, C., Fields, C., et Finney, J.L. 2020. Élaboration d'un cadre de modélisation de la répartition des espèces et son application à douze espèces sur la côte canadienne du Pacifique. Secr. can. de consult. sci. du MPO. Doc. de rech. 2020/004. xiv + 118 p.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Species distribution models (SDMs) have become increasingly valuable as a tool for the management and conservation of marine resources and places. However, despite their utility and prevalence, critical aspects of SDM development and uncertainty assessment are routinely overlooked. Best practices are therefore warranted and their application is increasingly required. This framework, implemented with purpose-built scripts written in the R statistical programming language, has been prepared as both a tool and a set of guidelines and methods for the development of consistent, interpretable, and defensible SDMs to support DFO's contribution to Canada's ocean policies. SDMs were built for twelve benthic species to illustrate the application of the framework, and guide emergency oil spill response planning as part of the Regional Response Plan for the Northern Shelf Bioregion. Three model building methods of increasing complexity were applied using a suite of the best available environmental predictors. Knowledge-based envelope models were produced for all species, and emphasized for those found to be data deficient. These envelope models provided guidance for the development of the subsequent data-driven models, and can be used to help evaluate uncertainty in model predictions. They also provide an avenue for engaging species experts in the process. Data-driven generalized linear (GLM) and boosted regression tree (BRT) models were generated, along with a corresponding ensemble model, for the eight species found to have adequate observational data. The highest quality predictions were generated for those species for which sample sizes were high, and observational data were well-distributed across the study area. All ensemble models performed moderately well (AUC > 0.7) when evaluated using a spatial block cross-validation approach. Thirteen recommendations were conceived as part of the development of the framework and its application. They provide guidance on the application of SDM methods related to data selection and preparation, model development and evaluation, and highlight ways to improve modelling outcomes by applying best-practices where possible. Future development of SDMs for additional species in Pacific Canada will be greatly facilitated by the set of common predictors, methods, and evaluation tools assembled here.

1 INTRODUCTION

Informed decisions about the management and conservation of marine species and their habitats increasingly rely on understanding their potential distributions. Since few marine species are inventoried, species distribution models (SDMs) have become a common approach to estimating distributions of valued species. SDMs are defined here as models that relate occurrence data with environmental conditions at the known occurrence locations (Elith and Graham 2009). By predicting a species' distribution based on correlations between observations and environmental predictors, SDMs provide a rapid, cost-effective way of estimating where species of interest are likely to occur. Understanding potential species' distributions can inform a variety of management activities including marine spatial planning, assessing fishing impacts, and invasive species control. Emerging applications include changes in species distribution in response to climate change, impacts of commercial fishing, stock assessment, regional harvest planning, and emergency pollution response. SDMs can also identify gaps in ecological knowledge, helping to target future survey and research efforts.

However, there are several challenges to building SDMs. These include the increasing accessibility of both sophisticated statistical methods and environmental data, the variety of species occurrence data, and the diversity of sampling methods. There are also numerous considerations related to model objectives, data preparation, variable selection, and analytical methods that influence the selection of appropriate methods and model interpretation. These considerations make identifying the best method for any particular application challenging (Gregr and Chan 2014; Yates et al. 2018).

The goals of this document are to standardize and facilitate the SDM development process by presenting a framework to guide data preparation, model fitting, model evaluation, and the interpretation of results and uncertainties. The framework is focused on SDMs built from species occurrence (presence-absence) data. Estimates of abundance are not considered. The framework is informed by current best practices, and its application is illustrated with a diverse suite of benthic species from Canada's Pacific coast. It is implemented with purpose-built scripts, written in the R statistical programming language. To reduce the need for definitions of terms within the document, a glossary of relevant terminology is included.

1.1 OVERVIEW OF SPECIES DISTRIBUTION MODELLING

Over the years, work on understanding how species are distributed across space and time has developed independently in a variety of disciplines, and various terms have been applied (Hirzel and Le Lay 2008). While this framework focuses on data-driven SDMs, it also considers the utility of envelope models, which provide a way of bounding species' distributions using presence-only data, or simply hypothesized environmental constraints. They are used here to describe the current ecological understanding of how species are related to their environment, and to evaluate how well this ecological understanding is reflected in the data-driven SDMs.

It is generally accepted that species distributions are a function of biotic interactions and barriers to dispersion in addition to environmental factors (Soberón and Peterson 2005). However, the literature remains equivocal on whether predictions based on the relationships between species observations and the environment reflect only suitable habitat or can be extended to the distribution of species. For example, Guisan et al. (2002) argue that all such models are SDMs because species observations are necessarily constrained by these three factors. Others (e.g., Araújo and Peterson 2012) have argued that despite the constraints on the observational data, important processes remain unaccounted for if only environmental predictors are used. Recent work (Dallas and Hastings 2018) has found that the abundance-suitability assumptions

underlying SDMs seem to depend on study area extent and resolution. Thus, the reality is likely that models based on environmental correlations likely fall somewhere on a continuum between habitat suitability and species distribution, depending on the ecology of the species, its exploitation history, the quality of the available data, and the scale of analysis.

Before building an SDM, it is useful to consider the overarching context of the desired model. Aspects of model context include the management need, what is understood about the species and its habitat, and the available data (Figure 1.1). A clear understanding of model context, in addition to the consideration of assumptions required at each stage in the model-development process, will help inform both the most appropriate modelling methods, and facilitate the interpretation of the model results.

Best practice includes the definition of an ecological model describing the context for the analysis, a data model describing the occurrence and predictor data, and a statistical model relating the two. As described by Austin (2002), the ecological model is comprised of the ecological knowledge and the model objective, which should reflect the management need. The data model includes how the occurrence and predictor data were collected, measured, and organized. The statistical model describes the relationships between the occurrence and predictor data. Accepted modelling practice begins with the ecological model, which aids in the construction of the data model. Together, these two then inform the data preparation steps (Figure 1.1), and thereby the selection, application, and interpretation of the statistical model. For example, understanding how occurrence data were sampled can constrain the life stage being modelled, which has implications for the data used, and what can be inferred from the model predictions.

SDMs can vary greatly in terms of complexity but are generally developed using a combination of ecological understanding and correlation analysis. Envelope models based entirely on ecological understanding can provide useful information at broad spatial extents (e.g., Kaschner et al. 2006), while more complex SDMs based primarily on correlations can provide reliable predictions over short time scales at local extents (e.g., Maxwell et al. 2015). Selecting appropriate statistical models will depend on what is known about the species ecology, the management objective, and the availability and applicability of the occurrence and predictor data.

It has generally been assumed that model predictions will be more transferable (i.e., provide reliable predictions in places or at times beyond the study area) if the models are less complex (Levins 1966; Yates et al. 2018). While there is some evidence to support this view (e.g., Gregr et al. 2018; Randin et al. 2006; Tuanmu et al. 2011), others have argued that complexity is not the problem so much as relationships fitted to biased data (Evans et al. 2013; García-Callejas and Araújo 2016). Nevertheless, there is agreement that models which emphasise ecological processes will transfer better to other times and places (e.g., Bell and Schlaepfer 2016; Yates et al. 2018). When model objectives require models to be transferable (e.g., potential range expansion or shift), validation with additional data collected at different times or locations should be attempted (Figure 1.1). Model evaluation with independently collected, well-sampled data is considered best practice for assessing a model's transferability (Araújo et al. 2019).

Modelling species and their distributions has come far from its systems ecology roots where mathematical descriptions of ecological systems were first developed (e.g., MacArthur 1955; Ord 1979; Wilson and MacArthur 1967). However, the practice of systematic, explicit, deliberate model design that has long been advocated (e.g., Levins 1966; Silvert 1981) remains best practice today. Fundamentally, good model design means understanding and explicitly resolving trade-offs between model accuracy, generality, and precision.

Due to increasing computing power, data availability, and on-going advances in modelling methods, SDMs can now be developed relatively simply. However there are many steps to model implementation (i.e., Figure 1.1), and each step (choice of methodology, data collection, data preparation, variable selection, and model selection) will introduce error and uncertainty into the final model predictions. Model reviews (Gregr and Chan 2014; Planque et al. 2011; Robinson et al. 2017) show that critical aspects of model design and development, and uncertainty assessment are routinely overlooked.

Given the ease with which models can now be developed, they are commonly built despite the potential unsuitability of available data to model objectives. Best practices are therefore warranted and their application is increasingly called for (Araújo et al. 2019; Grimm et al. 2014). Araújo et al. (2019) offer a comprehensive set of best practices addressing occurrence data, predictor data, model development, and model evaluation, and suggest a sliding scale ranging from deficient to (largely aspirational) gold standards. While not adhering strictly to these guidelines, this work pays close attention to each category, avoiding deficient practices, and excelling at others.

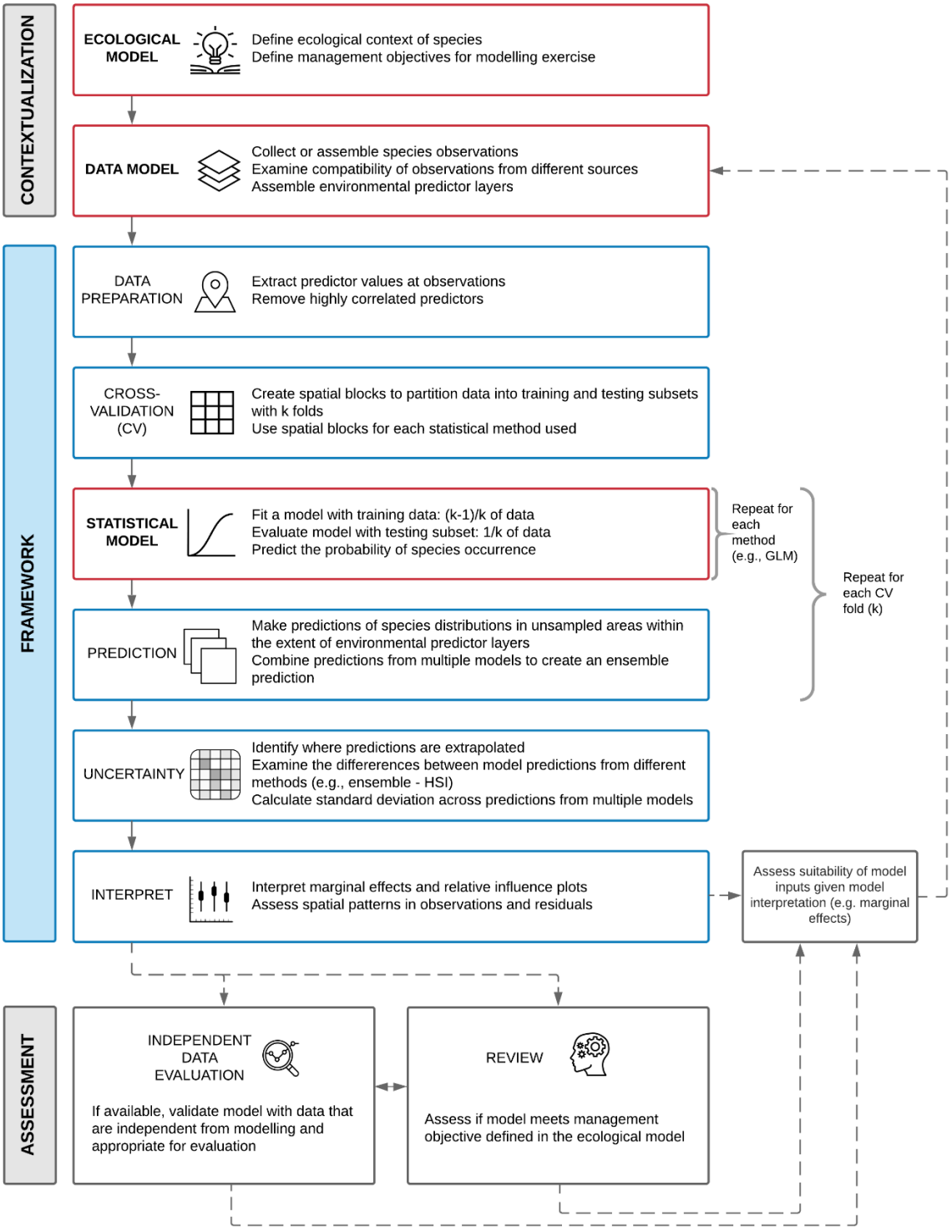

Figure 1.1. Overview of the complete modelling process from developing the model context (Contextualization) to the assessment of the model predictions. The Framework developed here for species distribution modelling includes a series of prescribed steps that automate best practices. Generalized linear models (GLMs) are an example of a data-driven modelling method, while habitat suitability index (HSI) models are an example of a knowledge-based envelope modelling approach.

1.1.1 The ecological model

The management of a species or its habitat is inseparable from its ecology. Consequently, the ecological model describes both the ecological context and the relevant management aspects. Existing understanding of the species' life history, environment, movement patterns, and exploitation levels provides the ecological context within which the model will be developed, and the assumptions that will need to underpin it. Management aspects provide the rationale for the modelling exercise (e.g., conservation, restoration, spatial planning), thereby informing the relevant extents and resolution. A well-defined ecological model allows for a clear modelling objective to be defined, and informs the selection of model extents and resolution. The context provided by the ecological model is thus important for defining the data model.

Species characteristics will also influence the accuracy of the resulting predictions (e.g., Elith et al. 2002; Garrison and Lupo 2002; Hepinstall et al. 2002). Such characteristics may include conservation status, habitat distinctiveness, local movement and migratory behaviour, range size, and trophic level. Studies exploring the impacts of these traits have had differing results (see McPherson and Jetz 2007), emphasizing that potential life history and biological interactions can influence model results. Understanding the assumptions necessary to represent such species characteristics, and the associated biases in the observational data is essential to defensible and reasonable inference.

1.1.2 The data model

The data model consists of species observations and environmental predictors (the variables that characterise the habitat of the species). Observations (i.e., the occurrence data) come with an entire context that describes how, when, why, and where they were collected. Similarly, the context for predictor variables (the independent data) includes their resolution, whether they are static (e.g., elevation) or assumed so (long term averages of observations), and any presumed interactions between them. Different questions will often require dynamic predictor variables to be scaled differently.

Occurrence data can range from presence-only observations with no information on where the species was absent, to abundance or density data obtained through systematic surveys. While systematic surveys are considered best practice (Araújo et al. 2019), they are rarely implemented because of limited resources. Therefore, to increase sample size, SDMs are often developed using assemblages of data from a diversity of observers, platforms, times, and places, often collected using different sampling methods. Other important metadata includes their temporal and spatial accuracy, and whether they are tied to particular life history stages (e.g., larval vs. adult) or associated with important seasonal events (e.g., maximal growing or spawning periods). This ecological information is critical to model interpretation, as observations are often aggregated across such ecological dimensions.

The availability of predictor variables and their resolution (both spatial and temporal) is the next practical consideration. This is a key determinant of the spatial extents, resolution, and ecological richness (sensu Gregr and Chan 2014) of the model. Environmental predictors fall on a spectrum of proximal to distal, depending on their position in the chain of ecological processes that link them to the species' distribution (Austin 2002). The use of proximate predictors tends to produce models that are more robust and generalizable, but such predictors can be difficult to collect (Austin 2002). Distal variables rely more on assumptions about process, making the validity of such assumptions central to the effectiveness of the model when forecasting in space and time.

Variable selection is a critical aspect of the data model, and is ideally informed by the intersection of model objectives and data availability. However, data availability is often the

single most important criterion. This makes it important to ensure that the available predictors are representative of both the area of interest and the presumed ecological processes. Understanding assumptions about representativity is critical for model interpretation and uncertainty assessment.

1.1.3 The statistical model

There are many ways to relate species to their environment. Methods are available for species with little or no data, presence-only data, presence-absence data, and abundance or density data. They range in complexity from simple envelope approaches that focus on the physiological limits of species, to an array of regression-based, and machine learning methods. One of the simplest approaches (habitat suitability index modelling, e.g., Kaschner et al. 2006) is based entirely on existing knowledge. Ecological niche factor analysis (Hirzel et al. 2002) and BIOCLIM (reviewed in Elith et al. 2006) are other envelope-type methods, most of which use presence-only data. When observations of presence or abundance (e.g., counts) and absence are available, a range of increasingly data hungry approaches are available. Classic regression methods (i.,e., generalized linear models, generalized additive models, generalized additive mixed models) can be used to predict functional forms of increasing complexity, while machine learning methods (e.g., classification and regression trees, neural networks, maximum entropy) can generate more complex relationships. The more complex the model, the more data are required to parameterise it. Given the diversity of methods and complexities, the principal consideration in selecting modelling methods is how well they meet the management objective of the ecological model. In practice, the availability and relevance of the observational and predictor data (i.e., the data model) are also primary considerations.

Studies comparing the relative performance of the various methods appear regularly (e.g., Elith et al. 2006; Guisan and Zimmermann 2000; Pearson et al. 2006; Segurado and Araújo 2004). While the studies differ in detail, flexible regression methods and some machine learning approaches tend to be best at explaining the observational data. However, simpler approaches (e.g., niche or envelope models) can be sufficient for some applications such as predicting potential range or range shifts across large spatial extents (Cheung et al. 2009), or supporting data poor contexts (Gregr et al. 2018).

1.1.4 Incorporation of local ecological knowledge

Descriptions of species distributions can benefit from the integration of local ecological knowledge (LEK) arising from regular, long-term contact with the species and its environment, including traditional ecological knowledge (TEK) passed down through generations. Ericksen and Woodley (2005) find that the relevance, credibility, and legitimacy of ecosystem models increase when the knowledge, needs, and concerns of those primarily affected by the modelling outcomes are taken into account. While methodological guidelines for the integration of LEK in models are still being developed, there are a growing number of studies attempting this integration (Belisle et al. 2018). To incorporate LEK, Belisle et al. (2018) recommend a reproducible and multidisciplinary approach that incorporates concepts from the social sciences and ecology to process empirical data and knowledge which may take the form of myths, legends, or rituals (e.g., Colding and Folke 2001).

While this framework and its application do not incorporate LEK into the SDMs, any future applications of the framework would benefit from its inclusion, when available. LEK can support the model development process at several stages, from the development of the ecological, data, and statistical models, to independent model validation. LEK can be directly incorporated into a knowledge-based envelope model during the expert elicitation process, and could be combined with empirical methods in a Bayesian framework (Belisle et al. 2018).

2 DATA SELECTION

This section describes the characteristics of species occurrence and predictor data that can lead to bias and the resulting introduction of prediction errors leading to poor model performance and interpretation. All data come with biases, and all models rely on assumptions about the appropriateness of the data for the model purpose. For example, observations are often collected along well-travelled routes or only within a portion of the true range or habitat of the species. The resulting spatial sampling bias can lead to the model describing the patterns at the sampled sites rather than across the species' range (Barry and Elith 2006). There can also be bias or error in the predictors from averaging, artefacts in the source data, or interpolation methods that can all lead to local prediction errors (Barry and Elith 2006). Best practice includes the consideration of such biases in occurrence data during model development and interpretation, a practice applied herein.

Occurrence data are typically obtained as points or polygons while predictor data are commonly represented as point grids or rasters. The relationships between occurrence and predictor data will depend on the extents and the spatial and temporal resolution of the analysis (which determines how the observations and predictors are averaged).

2.1 OCCURRENCE DATA

Data on species observations come in many forms, and are collected in different ways. The most common, and least powerful, observations are non-randomly collected observations of presence, with no credible absence data. Nevertheless, the abundance of such data sets has led to methods developed to facilitate the use of presence-only data in SDM development (e.g., Elith et al. 2006; Phillips and Elith 2010). This SDM framework focuses on observations collected via surveys designed for monitoring, stock assessment, or fisheries catch reporting. All these data sources contain (or allow the estimation of) absence observations, adding important information to the modelling process.

Often, observations will also record abundance in the form of counts, density estimates, or catch per unit effort (CPUE). While abundance data are essential to calibrate density models, they are also less widely available, and require more advanced model treatment. This advice therefore focuses on binary, presence-absence data to allow a broader array of data and species to be considered. Large parts of this advice are nevertheless relevant for abundance data.

In addition to the type of data collected, there are other characteristics of the occurrence data that can influence a model's ability to describe the species' spatial distribution, and discern the influence of predictors.

2.1.1 Sample size

The number of observations used for modelling will depend on a variety of factors because there is a trade-off between data quality and quantity. Data quality depends on the survey methods, the extents to which they were applied, and the length of the time series. While models can be constructed with fewer than 100 observations (Hernandez et al. 2006; Stockwell and Peterson 2002), a sample size of several hundred is more typical to effectively explore statistical modelling. Quality absence data, and consideration of the number of presence and absence observations (i.e., prevalence) will lead to better model outcomes (Barry and Elith 2006; Lobo et al. 2010).

Combining observations from multiple surveys to increase sample size should be considered when sample sizes of presence or absence observations are low or when one survey does not cover the modelling extent. However, combining surveys with different species detectabilities

may increase model uncertainty by introducing multiple sampling biases arising from these differences which may be difficult to disentangle.

2.1.2 Extents

The most powerful observational data will span the full spatial extent of the species' distribution and include a range of differently suitable habitats. This is equivalent to spanning the entire range of suitable environmental conditions. This informs the spatial and environmental extent to which model results can be applied. The temporal extents of the observations are also relevant to understanding the ecological context, including whether the data were collected over one or multiple years, or during different seasons. This informs how well the results generalise to other years, and the degree to which they reflect a particular time of year (Gregr 2011; Gregr et al. 2018). Accordingly, the temporal extents of occurrence data will influence the relationships with dynamic predictors (e.g., temperature, salinity), and may lead to a model fit to an anomalous set of observations. Further, if observations do not span the full extent of the species' range, important habitat descriptions may be missed (due to missing true presence data), or incorrectly discriminated (because of missing true absence data).

2.1.3 Sampling bias

Best practices require observations from surveys with designs that are statistically robust and unbiased, with observations distributed across the species' range. Non-systematic surveys can lead to predictions of patterns in the data collection methods rather than the habitat associations, especially if the sampling is spatially or environmentally structured (Araújo et al. 2019). For example, fisheries catch data typically occur in areas of high habitat suitability, since fishers have a strong incentive to avoid areas of low species' density, often leading to the assumption that avoided areas represent true absences. Observations can also have a temporal component with seasonal and annual changes in behaviour, mobility, or life history stages. If observations are limited to a particular life history stage (e.g., spawning) or time of year (e.g., migration), these limitations need to be explicitly considered in the model interpretation.

Lastly, the potential bias of local extirpation is an important consideration for exploited species. If a species has been removed from a particular part of their habitat, like shallow waters, or a particular region because of ease of harvest, and no observations exist for those areas, the habitat relationships may not be correctly captured using correlative models. This is particularly true if the extirpated range has unique habitat characteristics. In such cases, any resulting models will likely under-represent the species' potential habitat.

2.1.4 Precision

The precision with which observations are made and recorded includes spatial, temporal, and taxonomic dimensions. The spatial precision of the occurrence data should inform the spatial resolution of predictor variables, while their taxonomic precision is relevant to what can be inferred from the resulting SDM. Similarly, temporal precision is important when the management objective relates to particular life history stages or seasonal differences in habitat suitability.

Ignoring the precision of occurrence data and using all available data blindly is considered a deficient practice (Araújo et al. 2019). This is particularly true if observations are pooled from different collection methods, which can have variable precision across space, time, or

taxonomy. Observations can be made from vessels or underwater, remotely or directly, and all have consequences for how precisely an observation can be described.

2.2 PREDICTOR DATA

Habitat models rely on environmental predictors to describe the abiotic world. Relevant predictors of the marine environment are available from a variety of sources and can be represented as either continuous or categorical data. Remote sensing technologies (e.g., satellite imagery, drones and acoustics) and ocean models (including elevation, bottom type, and ocean dynamics) provide a source of increasingly detailed data for a growing number of environmental variables.

Satellite imagery produces models of sea surface temperature, chlorophyll-a concentration, euphotic depth, and wind speed and direction. Ocean circulation models estimate temperature, salinity, mixed layer depth, and current speed and direction at a range of depths. When coupled with a biogeochemical model, circulation models may also predict biological and chemical properties such as concentrations of dissolved oxygen, aragonite, and plankton. These dynamic predictors can be represented at different temporal scales (e.g., seasonal, annual, decadal). Bottom type, an important determinant of habitat for many species, can be interpolated from observations (Gregr et al. 2013) or predicted with models (Li et al. 2011). Finally, topography represents an important class of commonly used predictors that includes depth and a variety of derivatives such as slope, curvature, bathymetric position index (Walbridge et al. 2018) and measures of rugosity (e.g., Du Preez 2015; Sappington et al. 2007).

Spatial variables, such as distance to important physical or biotic features (e.g., shoreline, reefs, or kelp forests) can also be included. Some models have included latitude and longitude as proxies for unknown, unrepresented, or unrepresentable predictors, however the interpretation of such geographic variables is problematic. Additionally, fishing effort can be an important predictor of species distributions in impacted ecosystems (e.g., Foster et al. 2015; Tien et al. 2017).

2.2.1 Extents and resolution

The range of values in the predictor variables used will typically change in response to the size of the study area (e.g., consider the depth range covered in a coastal vs. a regional model). Such differences in extents can affect the measured relationships with species observations (Austin 2007). At smaller extents (i.e., within a species' range) a species' preferences may be uncovered, while larger extents may identify the biological limits of a species. When the range of a predictor far exceeds the species' habitat suitability and absence data are unavailable, the biological relevance of the relationships can be compromised (Austin 2007; Fourcade et al. 2018).

The resolution of available marine predictors can span orders of magnitude. Remotely sensed data and circulation models can range from the kilometre to the metre scale, while acoustically-derived topographic predictors can be resolved to sub-metre scale. For computational reasons, resolution is invariably tied to extents (larger study areas will require reduced resolution), and reduced resolution leads to reduced precision as predictor variables are averaged over space. To best predict a species' habitat, the resolution should be relevant to how the species interacts with its environment (Wiens 1989). For example, the spatial resolution of relevant predictors will be very different for an intertidal snail compared to a highly migratory cetacean. Similarly, because habitat processes operate at multiple resolutions (Levin 1992), predictors aggregated at daily, monthly or seasonal resolutions may help capture additional important information compared to long-term climatologies. It is therefore important to consider how well predictors

match the resolution of the observations both spatially and temporally (Wiens 1989). For example, rugosity derived from a 100 metre resolution depth layer may not be meaningful for species with a known relationship with rugosity at a sub-metre scale.

Resolution also influences how data are prepared, and the interpretation of results at local scales. For example, if a grid cell is not sufficiently sampled, a false absence may result. Similarly, while a global study may predict high suitability for a 10 km x 10 km region of the coast, real-world heterogeneity suggests it is unlikely the suitability of such a large area will be uniform. This has implications for model testing and interpretation.

2.2.2 Accuracy

While physical predictor data collected concurrently with observations would be the most reliable, modelled predictors allow SDM predictions to be interpolated across the entire area of interest. Modelled predictors inevitably misrepresent some aspects of the real features leading to relative (e.g., tidal fronts where freshwater inputs are not included, or features smaller than the model resolution), and absolute (e.g., sandy sediment predicted instead of mud) representational errors. If such misrepresentation is systematic, it can be identified by examining the predictor's derivative (e.g., curvature can highlight artefacts in the slope predictor). Temporal averaging or scaling can cause similar misrepresentations (see following section). However, all models will have error, and even with strong validation tests, the spatial distribution of these errors is difficult to find. While best practices recommend assessing how uncertainties in the predictors influence model results (Araújo et al. 2019), this is acknowledged to be a significant challenge requiring sophisticated simulation methods. Such simulations are beyond the scope of this work. Instead, the uncertainties in the accuracy of predictor variables are acknowledged and their possible effects on model interpretation considered.

2.2.3 Scaling

Predictor variables are available at a variety of spatial resolutions, requiring some to be scaled to the chosen model resolution. Standard practice is to average those at a finer resolution than the model resolution, while coarser predictors can be interpolated, resampled, or downscaled (i.e., the same value is applied to all cells within the each coarser study unit). These spatial scaling considerations are sufficient for static predictors (i.e., those assumed to not vary over time such as bathymetry and its derivatives).

Most SDMs in the literature continue to be presented as a single static, average-conditions map. This means that any predictors with a temporal dimension (e.g., dynamic predictors such as temperature, salinity, and chlorophyll-a) need to be averaged across time. Best practice requires these averaged values to reflect the biological response being modelled (Araújo et al. 2019), because how a predictor is averaged may influence its relationship with the species observations (Eger et al. 2016; Levin 1992; Wiens 1989).

Thus, predictors are often averaged across a time series for a period or duration relevant to (the ecological or management aspects of) the ecological context. Ideally the duration will also overlap with when the species observations were collected. The temporal resolution and extent of the model can be important to management either seasonally (e.g., spawning habitat may be of interest, or summer occupancy for migrating species) or across years (e.g., protected area design, response planning, or fisheries management). Dynamic ocean management (Maxwell et al. 2015) provides an example of distributions models with high temporal resolution. Higher resolution models could benefit from the inclusion of temporally stochastic events such as oxygen dead zones, salinity dips, or lethal events such as ship strikes or contamination events.

The inevitable mismatch in resolution between observations and predictors (and often among predictors) means that some correlations will represent an ecological process well, while others will serve as proxies for other variables or other resolutions. Austin (2002) characterized this as a distinction between proximate and distal variables, although in reality this is likely more of a spectrum of association between observations and predictors. How important this distinction is will depend on the model objectives. If the objectives relate to understanding process, or model transfer, proximate variables are desirable. If explaining the observed pattern within the study area is the primary objective, then distal variables may be equally suitable, even though they may be proxies for other variables or processes.

2.2.4 Spatial predictors

Spatial autocorrelation (SAC) can play an important role in species distributions. Positive SAC is often observed in species data whether observations are binary (presence/absence) or continuous (abundance). The spatial dependence among observations can stem from different sources, for example, biotic factors such as dispersal or the presence of SAC in the physical conditions such as temperature (Bahn and McGill 2013). Including a measure of spatial structure has been shown to improve model performance (Augustin et al. 1996; Bahn and McGill 2007; Martin et al. 2014). There are several approaches for incorporating information on spatial structure into SDMs (see Dormann et al. 2007; Elith and Leathwick 2009); however, the merits of the different approaches remain a topic of debate (Bini et al. 2009; Hawkins 2012; Kühn and Dormann 2012). They range in complexity from simple approaches such as including geographic coordinates or an auto-covariate as predictor variables to more complex methods using covariance functions to approximate spatial random fields. With any approach, it is important to consider the scale of SAC relative to the spacing of your samples if using spatial models for prediction as gaps between samples larger than the scale of SAC will lead to poor predictions from spatial models.

3 A SPECIES DISTRIBUTION MODELLING FRAMEWORK

This section describes the SDM framework, developed in consideration of best practices for model building, and refined through experience obtained during the framework application process. The framework (Figure 1.1) is comprised of six components: data preparation, cross-validation, model fitting and evaluation, prediction, uncertainty and interpretation. It was developed in conjunction with purpose-built scripts written in the R statistical programming language (R Core Team 2018). The framework workflow (Figure 3.1) can be repeated multiple times with different model fitting methods to create an ensemble model prediction (described in Section 3.5). Multi-model ensembles produce robust model predictions (Oppel et al. 2011) and provide insight into model uncertainty (see Section 3.6).

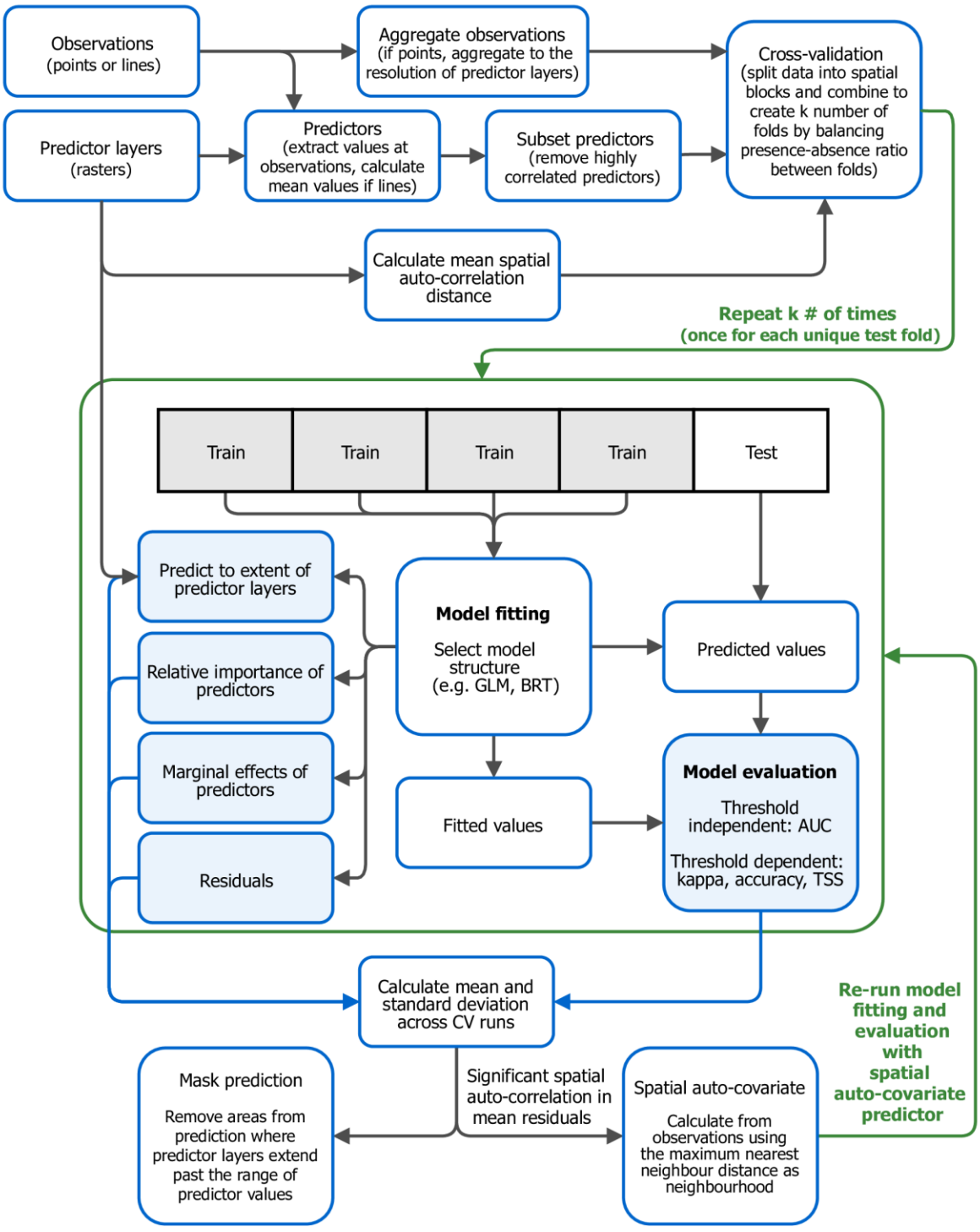

Figure 3.1. Diagram of the framework workflow implemented with purpose-built scripts. The workflow includes the processing of observations (dependent data) and predictor layers as well as model fitting, evaluation and prediction using a cross-validation approach. The model fitting approach varies with the choice of modelling method. Generalized linear models (GLM) and boosted regression tree (BRT) models are two examples of model fitting methods. Blue filled boxes represent outputs from the modelling process.

3.1 DATA PREPARATION

Once occurrence data and predictor layers have been obtained and a spatial resolution selected, there are several steps to preparing the data (given the considerations detailed in Section 2) prior to model fitting. First, observations and predictors must be represented spatially with the same coordinate reference system. Occurrence data are projected to match the geographic projection of the predictors. The observations are then checked for any erroneous (e.g., presence-absence values outside of 0 or 1) or missing values, which are removed.

Generally, observations are available as point locations (e.g., grabs) or line segments (e.g., trawl path). Predictor values are extracted from raster layers at the location of the observations, but the process varies slightly for point and line data. For point data, if more than one observation falls within a predictor raster cell, the points are aggregated to the raster grid. This standardizes the observations and predictors to the same spatial resolution. Additionally, by integrating observations from the same space over a number of years, observations are better aligned temporally with predictor layers that represent long-term average environmental conditions. For presence-absence observations, aggregation is completed by assigning presence to the cell if there is at least one presence observation within the cell (following Guinotte and Davies 2014). With presence/absence data, this step can lead to data biased towards presence. It is therefore important to consider whether modelling probability of occurrence using this aggregation method aligns with models objectives (e.g., modelling the probability of presence occurring at any time in the past). After observations are aggregated, the resulting change in sample size is calculated. If the number of observations has markedly changed (e.g., 50% reduction), a mismatch between the resolution of the observations and the environmental variables is evident, indicating the chosen resolution may not be appropriate for modelling.

For observations represented as lines, mean values are calculated for each raster cell that intersects the line segments (following Carrasquilla-Henao et al. 2018). For categorical predictor variables, the most dominant category occurring in the intersected raster cells is used. This approach integrates the predictor values for the entire line segment rather than just extracting a single value from the start, midpoint, or end of the line segment. Once observations are prepared for modelling they are mapped and can be visually examined for georeferencing errors (Figure 3.2).

Once the predictor values have been extracted to the observation locations from the raster layers, several data quality checks are required. Missing values in the predictor dataset, which can occur when observations fall outside the extent of any of the raster layers, are removed. The final number of records in the datasets now represents the sample size of the complete set of observations and predictors used for modelling. The predictors are then examined for model suitability by comparing the distribution of all values in the raster layers to the extracted predictor values. Ideally, the range of the extracted values should match that of the source raster that will later be used for prediction. If the range is truncated, model predictions in some areas will be extrapolated. Maps are produced for each predictor layer to allow for a visual check for any apparent artefacts on unexpected values. Variance inflation factor (VIF) and Spearman's correlation are used to assess collinearity among the extracted predictor variables. VIF values greater than 10 can indicate a predictor is highly correlated with at least one other predictor and should be considered for removal (Dormann et al. 2013). However it may not always be required or advisable to remove correlated predictors. Even if a highly correlated predictor is removed prior to modelling, determining the influence of its correlated predictors in the model is a challenge because one cannot disentangle their relative effects on the dependent variable.

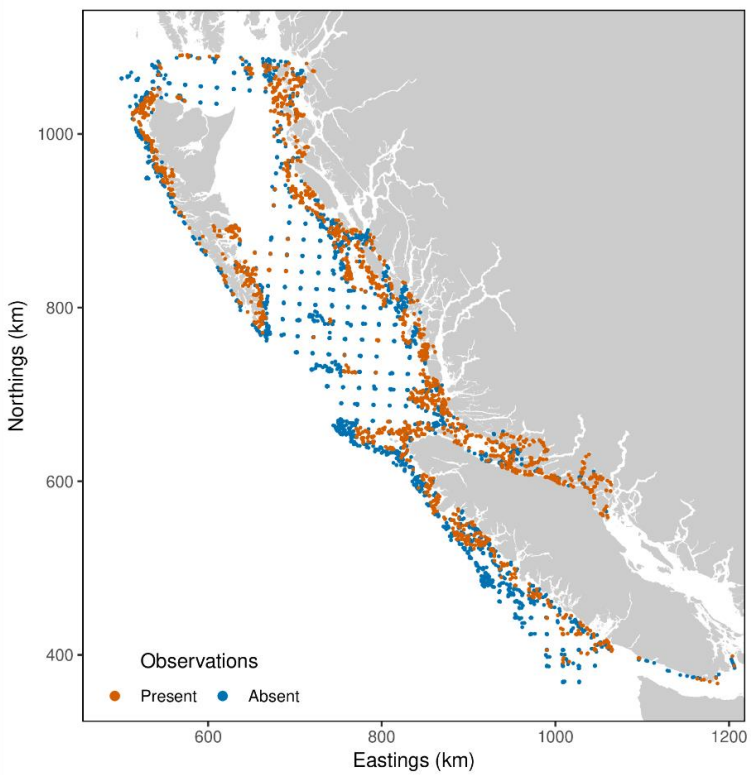

Figure 3.2. Distribution of presence and absence observations of an example species within a study area.

3.2 CROSS-VALIDATION

After the data model has been prepared, it is partitioned into training and testing datasets, used for model fitting and validation, respectively. Separating testing from training data is essential for evaluating model predictive performance (Hijmans 2012). Cross-validation (CV) is a widely used procedure (Roberts et al. 2016) that separates training and testing data in a way that allows all the data to be used in the model fitting process (Wenger and Olden 2012). CV works by dividing the available data into k folds. For each CV run, a different fold is reserved for testing and the remaining folds are used for training. The process is repeated k times until each fold has been used once for testing. Thus, for five-fold CV, five models are built, each with a different $4/5^{th}$ of the data and tested against the remaining $1/5^{th}$. Folds can be partitioned randomly or by blocking. Blocking can be spatial or according to other factors known to limit the independence of the data (see Roberts et al. 2016 for examples). Random CV continues to be used to evaluate SDMs despite being known to inflate performance scores because the partitions do not increase independence (Roberts et al. 2016). Independence of the training and testing data is a fundamental assumption underlying statistical tests of model performance (Legendre 1993). Spatial blocking (recommended and implemented in this framework) improves the spatial independence of the training and testing data, allowing for more accurate estimates of model performance and transferability (Fourcade et al. 2018; Hijmans 2012; Merow et al. 2014; Roberts et al. 2016; Trachsel and Telford 2016).

Using spatial block CV to increase the independence between folds is considered best practice when sub-sampling observations for model fitting and evaluation. This framework applies the spatial blocking procedure developed by Valavi et al. (2018) where the optimal block size is informed by the range of spatial autocorrelation in the predictor layers. The blocks are randomly assigned to folds, with several blocks within each fold. The process is repeated for 250

iterations and the folds with the most evenly dispersed sample size of presence and absence observations are used (Figure 3.3).

Model fitting and validation are performed multiple times, once for each CV run. There is no resulting single best model; rather, multiple best models are selected, each fit to the training data from their respective CV runs. Thus, the results (i.e., predictions, validation metrics, relative influence of predictors and marginal effects) are averaged and the standard deviation or range is calculated to illustrate the variation among CV models.

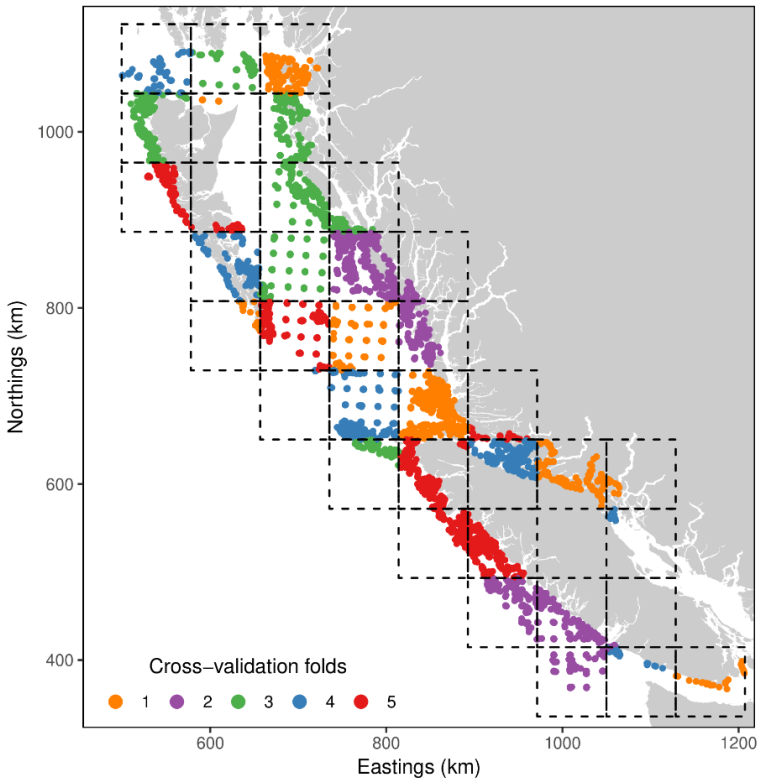

Figure 3.3. Spatial orientation of five-fold spatial block cross-validation of an example species within a study area. Each fold would be used once as testing data to evaluate models build with the training data made up of the remaining folds.

3.3 MODEL FITTING

Fitting a model to relate environmental predictors to species observations is the central technical exercise in SDM development. Model fitting can range from simple knowledge-based envelope model approaches (e.g., HSI) that do not require species occurrence data and rely on published relationships and expert consultation, to complex statistical algorithms (e.g., BRT) that discern correlations between predictors and species data. This framework considers fitting to include both model structure (selecting the appropriate predictors and their form) and model parameterization (identifying the parameters that allow the model structure to best explain the pattern in the observations). For knowledge-based approaches, model structure can take the form of linear, threshold, optima, exponential or logistic forms (Figure 3.4) and parameterization can be stepwise and iterative. When using data-driven, correlative methods (e.g., GLMs or BRTs), the structure and parameterization occur simultaneously.

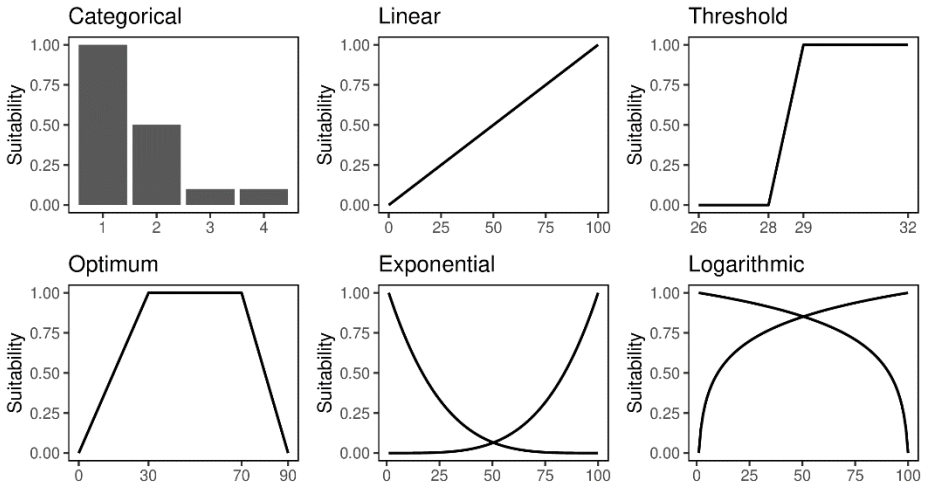

Figure 3.4. Examples of possible relationships between environmental predictor variables and habitat suitability used to develop habitat suitability index models.

The goal of model fitting is to explain as much of the variance in the observations as possible, without over-fitting the model. A model is overfit when its complexity (e.g., number of parameters) is greater than needed to accurately represent the relationship between the observations and predictors. Such a model will correspond too closely to the training observations and will perform poorly when applied to testing observations. This is understood to result in a loss of generality, leading to potentially poor model transferability (Wenger and Olden 2012). Knowledge-based envelope models cannot technically be overfit, as they use no data. They can, however, be over-specified for a particular area given locally derived expert opinion.

Modelling approaches use various statistics to balance under and over-fitting (Elith and Leathwick 2009; Merow et al. 2014). For regression type models (e.g., GLMs), Akaike information criterion (AIC) is commonly used as a quantitative, non-subjective method for selecting the best model from a number of candidate models by penalizing complexity in model fitting (Burnham and Anderson 2004). In this framework we use the Bayesian information criterion (BIC) to select the best model because BIC has a greater penalty for complexity than AIC (Galipaud et al. 2014; Link and Barker 2006; Schwarz 1978), resulting in simpler models. Another method of model fitting (not used in this framework) is regularization. Like BIC, regularization penalizes complexity but does so by reducing coefficient values, resulting in the coefficients of less influential predictors falling close to zero (Elith et al. 2011). For machine learning methods (e.g., BRTs), a portion of the data can be held out from model fitting to be used to calculate a stopping condition (Hastie et al. 2009). For BRTs this is achieved by iteratively evaluating the performance of the model using the holdout data at regular intervals during model fitting. As model complexity increases, performance measured against the holdout data will begin to decrease and a stopping condition will be triggered.

It is crucial that testing data remain independent of model fitting to obtain accurate model performance metrics (Araújo et al. 2019; Hawkins et al. 2003; Roberts et al. 2016). Therefore, this framework does not recommend investigating univariate relationships between the observations and each predictor prior to the separation of testing from training data. That step is often undertaken to help decide which predictors to include in the model; however, it violates the assumption of independence between the testing data and model fitting and can lead to inflated model performance measures (Hastie et al. 2009; Hawkins et al. 2003) and overfit models (Burnham and Anderson 2004). Instead, this framework applies variable selection in two stages. First, predictors without a known or theorized ecological relationship to the species being

modelled are filtered out prior to modelling. Secondly for correlative models, predictors that poorly describe the distribution of the observations are filtered out during the model fitting procedure. Predictors can be filtered out either by being removed from the model structure completely (e.g., during GLM model selection with BIC) or by remaining in the model structure but contributing little to the overall variance explained (e.g., BRT).

3.4 MODEL EVALUATION

SDMs can be evaluated for model fit and predictive performance. Model fit is assessed by measuring the variance remaining in the data used to build the model, while predictive performance is measured by examining the agreement between model predictions and observations not used in model fitting. Comparing statistics measuring model fit and predictive performance can provide a qualitative estimate of over-fitting. Higher performance values for the training over the testing data are indicative of an overfit model (Rooper et al. 2017; Wenger and Olden 2012). Evaluating model performance with data independent of model fitting is a necessary and important step in any SDM exercise, whether the goal is prediction or of an explanatory nature (e.g., building hypotheses regarding the environmental drivers of species distributions).

A variety of metrics exist to measure performance, including threshold independent and dependent statistics. Threshold-independent statistics compute the relationship between the observations and continuous predictions (e.g., probability of occurrence) and are recommended as the primary model performance metric for evaluating continuous model predictions (Lawson et al. 2014). Threshold-dependent statistics require continuous predictions be converted into presence-absence predictions using a specified threshold value. These statistics are recommended when presence-absence predictions are the end goal.

This framework adopts the popular area under the receiver operator characteristic curve (AUC) metric as a threshold-independent performance measure (Elith et al. 2006). The AUC ranges from 0 to 1. Values less than 0.5 indicate models that are worse than random, values of 0.5 indicate that the model is no better than random, and values of 1 indicate that the model perfectly predicts the occurrence data (Freeman and Moisen 2008; Merckx et al. 2011). According to Pearce and Ferrier (2000) and Jones et al. (2010) values of AUC greater than 0.9 are considered good, between 0.7 and 0.9 moderate, and less than 0.7 poor.

For comparison, because the utility of the AUC statistic has been questioned in the literature (Lobo et al. 2008), several threshold-dependent metrics are also calculated: sensitivity, specificity, true skill statistic (TSS), kappa and accuracy. TSS balances sensitivity (proportion of presence observations that are correctly classified) and specificity (proportion of absence observations that are correctly classified) and is independent of the prevalence of the observations (Allouche et al. 2006). Kappa is a measure of agreement between observed and predicted values that accounts for chance agreements and is dependent on prevalence of the observations. TSS and Kappa range from -1 to 1 with values less than 0 representing models that are no better than random and values of 1 indicating perfect agreement (Allouche et al. 2006). Values of TSS greater than 0.6 are considered good, between 0.2 and 0.6 moderate, and less than 0.2 poor (Jones et al. 2010; Landis and Koch 1977). Accuracy is the percent of predictions which are correctly classified and varies from values of 0 to 1 where 1 is the highest accuracy.

To calculate threshold-dependent metrics, this framework uses recommended best practice and determines an optimal threshold by evaluating model performance using a statistic such as TSS across multiple thresholds, rather than using a default of 0.5 (Freeman and Moisen 2008). The ideal statistic (e.g., TSS, Kappa or Accuracy) for threshold optimization will depend on the purpose for modelling, for example, a statistic that favours specificity over sensitivity may be

appropriate for certain management applications that require a low false positive rate (Freeman and Moisen 2008).

Several validation plots are produced using both threshold-dependent and threshold-independent statistics. These include receiver operator characteristic (ROC) curves on which the AUC is based (Figure 3.5), curves that show threshold-dependent statistics calculated across a range of threshold values (Figure 3.6), and a graphical contingency matrix showing the density distribution of true and false positive predictions, and true and false negative predictions (Figure 3.7). Examining such plots, though qualitative, is an important aspect of model evaluation. Validation plots can be more informative than a single statistic because statistics can be examined across a range of thresholds, and the density of observations that fall above and below presence-absence thresholds can be visualized. Validation plots can also be used to perform quality checks by ensuring curves match their related statistics (e.g., ROC curve matches the AUC statistic).

Lastly, residual validation plots are produced as a scatter plot of the residuals against each predictor variable. These plots can help inform how appropriate the model structure is for the data: if a strong residual pattern is observed, unexplained variation remains that may be better resolved with a different model structure.

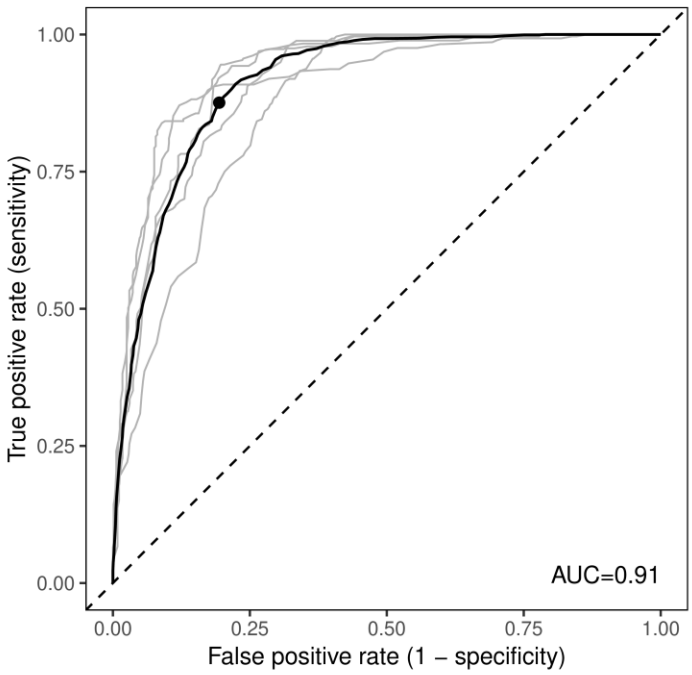

Figure 3.5. Receiver operator characteristic (ROC) curve. Grey lines represent ROC curves from five-fold cross-validation models and the black line represents the mean of the ROC curves. Curves are built from testing data. AUC is the area under the ROC curve. The black point denotes the false positive and true positive rate at the optimized threshold.

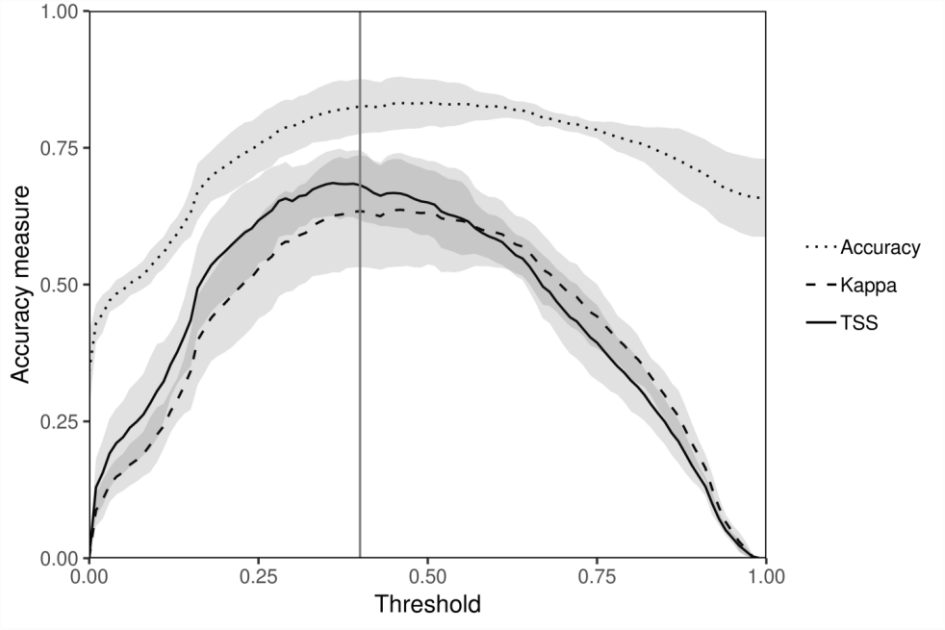

Figure 3.6. Curves representing threshold-dependent model accuracy measures across a range of thresholds. Curves are built from testing data. Grey ribbons represent one standard deviation around the mean calculated from five-fold CV models. TSS is the true skill statistic.

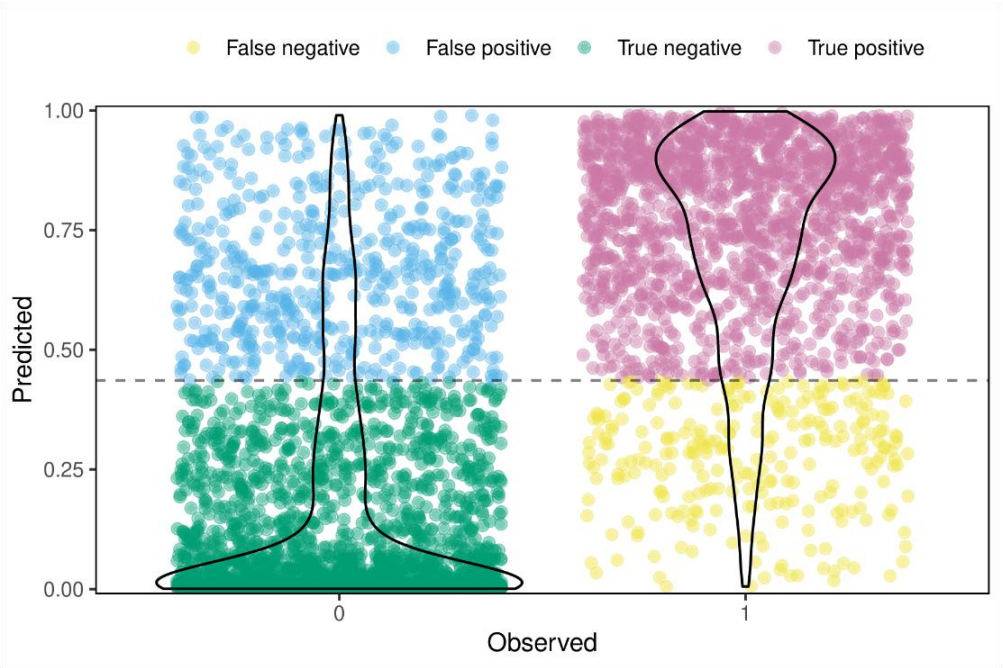

Figure 3.7. Visual representation of a contingency matrix showing the density of true and false positive predictions above the threshold and true and false negative predictions below the threshold. The threshold is represented by a grey dashed line. Black polygons are kernel density estimations which illustrate the density of predictions corresponding with present and absent observations.

3.5 PREDICTION

Once a model has been evaluated, the goal of SDMs is typically to predict to unsampled areas within the study area where predictor values are known. Within this framework, predictions are produced from models built from each CV run (following the workflow detailed in Figure 3.1). This workflow can then be repeated with different model fitting methods allowing an ensemble model prediction to be built for each CV run (Figure 3.8). Predictions are combined via a weighted mean using AUC for weighting (Anderson et al. 2016; Oppel et al. 2011). Prior to weighting, AUC values between 0.5 and 1 are rescaled to range from 0 to 1, with AUC values less than 0.5 set to 0. The modified AUC weights ensure that a model performing worse or no better than random (AUC <= 0.5) does not contribute to the ensemble prediction. This criteria for model entry into the ensemble was selected since any greater value (e.g., 0.7) would be arbitrary. Subsequently, the mean of the ensemble predictions is calculated and used as the final ensemble prediction (Figure 3.8).

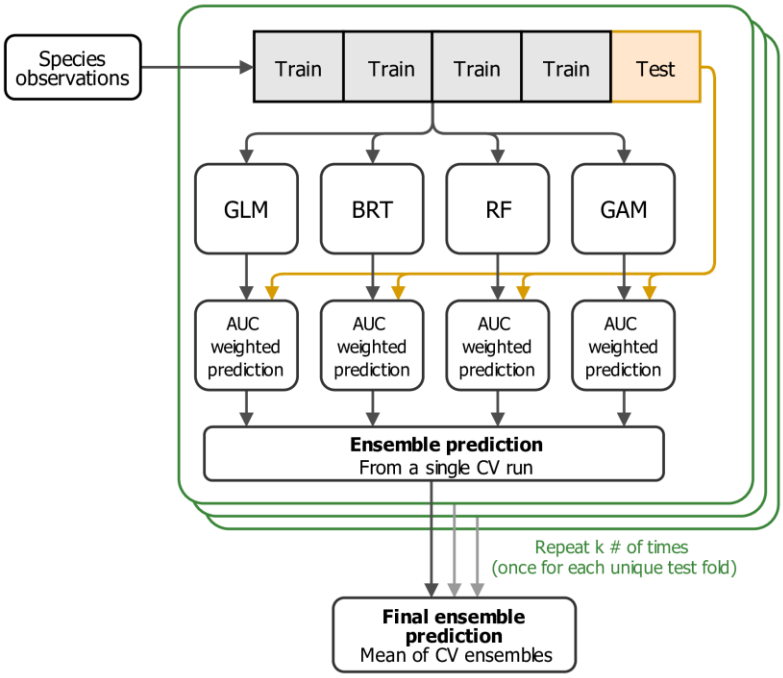

Figure 3.8. Diagram of the ensemble model workflow showing how predictions from multiple modelling methods can be combined into an ensemble within the cross-validation structure. Each model, including the ensemble, is validated with an identical set of testing data for each cross-validation procedure. Generalized linear models (GLM), boosted regression tree (BRT), random forest (RF) and generalized additive models (GAM) are four examples of model fitting methods.

In addition to the ensemble prediction, this framework recommends producing a knowledge-based model from available expert knowledge on the relationships between a species and environmental conditions. Such envelope models serve to illustrate how well expert knowledge around process can be represented, and provide a benchmark against which the data-driven ensemble model prediction can be compared (Figure 3.9).

Whether it is suitable to integrate envelope models into the ensemble, as opposed to using it as a comparative benchmark, depends on the considerations discussed in Section 1.1 such as whether model objectives align. An important best practice here is to ensure that when model predictions are combined into an ensemble, they represent the same units. For example, it would not be appropriate to combine predictions of abundance with predictions of probability of

occurrence. This framework does not include envelope models in the ensemble, but does support their use to explore aspects of model uncertainty (see Section 3.6).

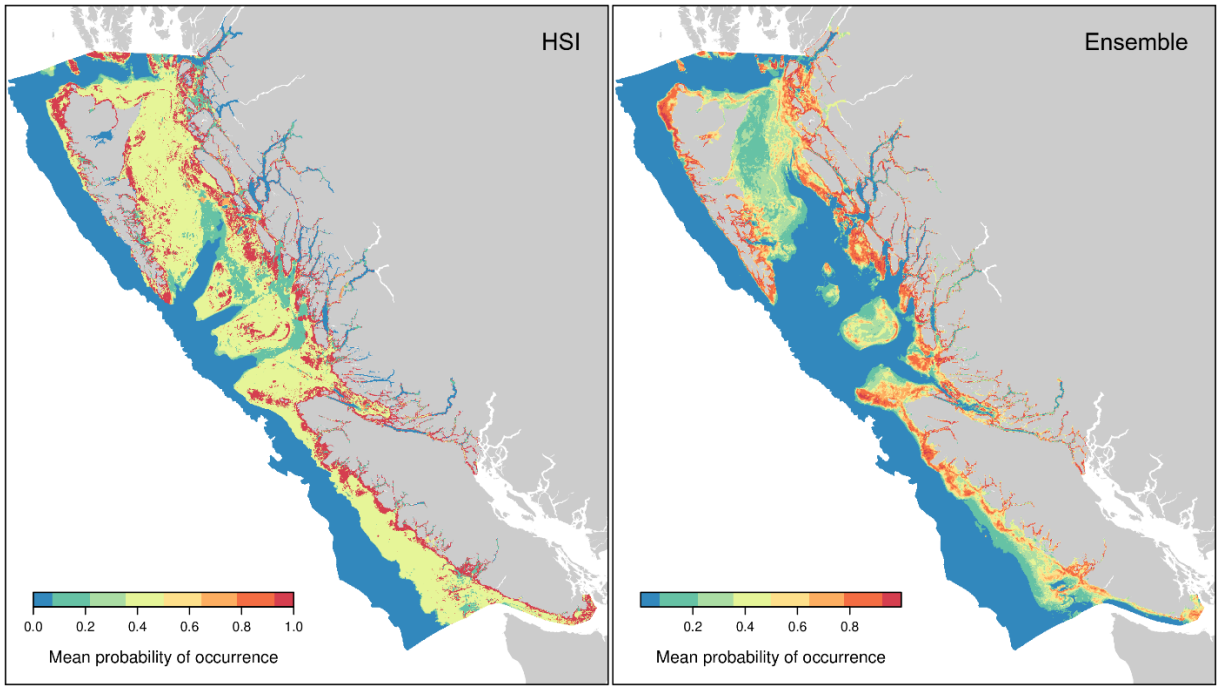

Figure 3.9. Probability of occurrence predictions from a knowledge-based envelope model (e.g., HSI, left) and a data-driven ensemble model (right). The ensemble prediction is the mean of the predictions from five-fold cross-validation models developed from multiple methods (e.g., GLM and BRT).

3.6 UNCERTAINTY

Once models have been used to predict to unsampled areas in the study area, the next step is to consider how uncertain those predictions are. Model uncertainty can stem from many sources such as the quality and availability of expert-knowledge, occurrence and environmental data, spatial and temporal sampling biases, parameter and model structure estimation, the statistical methods used, and natural variability in the system (Barry and Elith 2006; Dormann et al. 2008; Iturbide et al. 2018; Link et al. 2012). Estimating prediction uncertainty is essential to effectively inform policy decisions such as fisheries closures or marine spatial planning (Jones-Farrand et al. 2011; Jones and Cheung 2015), although uncertainty in model predictions is rarely assessed in marine applications of SDMs (Robinson et al. 2017). This framework presents four methods for reducing or describing uncertainty spatially across the study area.

First, uncertainty is constrained by identifying or removing areas of extrapolation (Figure 3.10). Extrapolation should be avoided in both predictor and geographic space (Austin 2007; Merow et al. 2014). Extrapolation in predictor space occurs when predictions are made in areas outside the range of predictor values used for model fitting for influential predictor variables (i.e., those that represent 95% of the cumulative relative influence, see Section 3.7). Extrapolation in geographic space occurs when predictions are made in areas outside the area covered by observations which can be represented by a minimum convex hull polygon around the observations. A well-performing model, developed using a spatial block CV approach, is appropriate for interpolation (i.e., predictions not extrapolated in geographic or predictor space). When predictions are extrapolated the performance of those predictions are unknown. The suitability of a model for extrapolation purposes can depend on the modelling method used. For example, GLMs produce functional predictor-response relationships that extend past the range

of predictor values in the training data while regression tree models extrapolate in predictor space using a constant value –the prediction at the minimum or maximum of the range (Elith and Graham 2009).

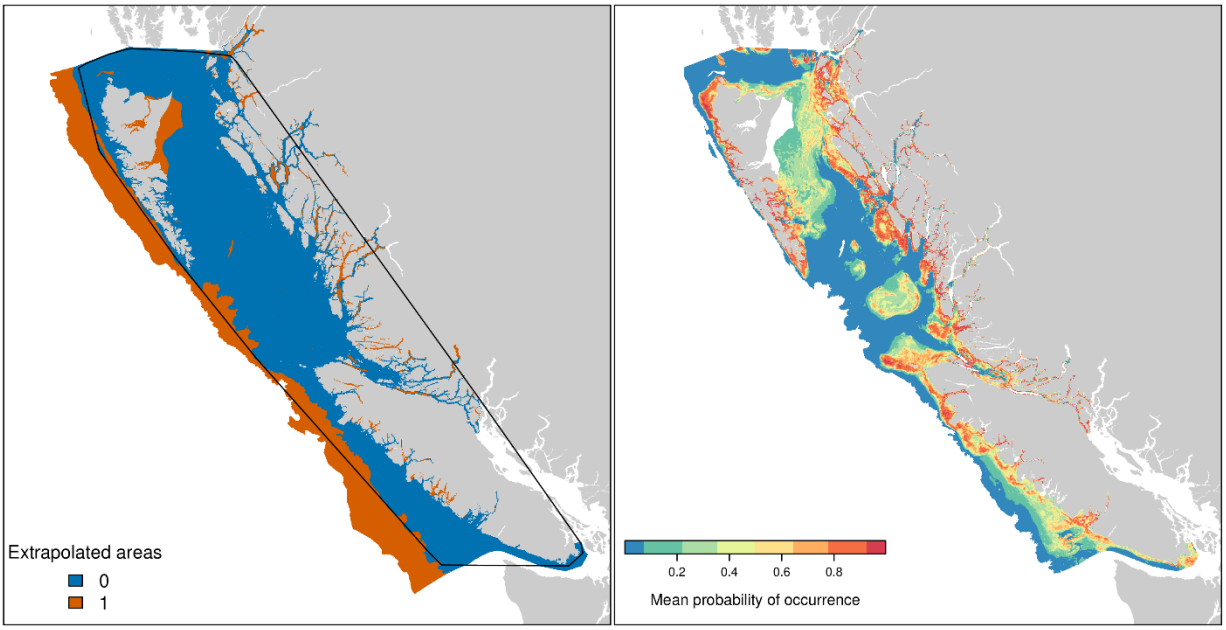

Figure 3.10. Extrapolated areas where predictors were outside the range of values used for model fitting (left). Mean probability of occurrence predictions from five-fold cross-validation models excluding the extrapolated areas in predictor space (right). The black line represents a minimum convex polygon around the observations.

Second, uncertainty is explored by contrasting the predictions from a knowledge-based envelope model with a data-driven ensemble model (Figure 3.11). Areas where the envelope and ensemble model predictions differ greatly represent areas where the probability of occurrence predictions are most uncertain. Without additional model validation and ground-truthing it is impossible to ascertain which prediction is more accurate in any given area. However, one can hypothesize what drives the differences between our current understanding of a species-environment relationship and relationships derived from correlative models.

Third, uncertainty in the ensemble model can be estimated by calculating the variation across the individual, component model predictions (e.g., those built with different model fitting methods). As recommended best practice is to incorporate multiple sources of error into a single estimate (Araújo et al. 2019), this framework incorporates both CV uncertainty and methodological uncertainty across the component models into a single estimate, but does not include parameter uncertainty. To estimate methodological uncertainty, the variance across the multiple model predictions that make up the ensemble (those built with different model fitting methods e.g., GLM, BRT) is calculated for each CV run and averaged (Araújo and New 2007; Jones-Farrand et al. 2011). To estimate CV uncertainty, the variance across the ensemble models from each CV run is calculated. The two variance estimates are incorporated into a single uncertainty metric by averaging the methodological and CV variance estimates and returning the square root to convert variance into standard deviation (Figure 3.12). Standard deviation was chosen instead of the coefficient of variation because it better highlights areas of high uncertainty where the probability of occurrence is also high and was considered more easily interpretable. This approach to uncertainty estimation can be applied to any modelling method where multiple model predictions are produced. For knowledge-based envelope

models, this would require building multiple model predictions from a variety of experts or knowledge sources.

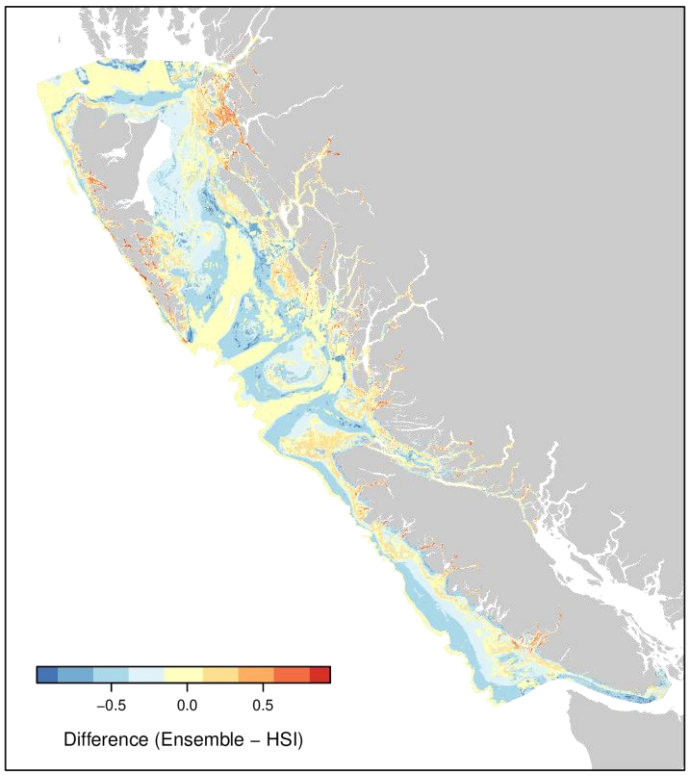

Figure 3.11. The difference between predictions from a knowledge-based envelope model and a data-driven ensemble model. Positive values represent areas where the ensemble model were predicting greater probability of occurrence than the envelope model.

Lastly, uncertainty is examined by mapping the mean residuals from the CV models for each model fitting method (Figure 3.13). Mean residuals from CV models are aggregated to a coarser scale (100 times the resolution of the predictor raster layers) and mapped to visualize any spatial pattern remaining in observations after model fitting. Areas with higher residuals indicate areas where predictions deviate further from the observations on average and thus are more uncertain than areas with residuals close to zero.

These assessments provide qualitative and quantitative estimates of confidence in model predictions as well as insights into aspects of data quality and sampling bias (through the comparison of knowledge-based and data-driven model predictions). The mapping of residuals also provides some insight into the variability (i.e., non-stationarity) of the species' habitat relationships. Uncertainty stemming from a lack of spatial precision in the observations, errors in the predictor layers, missing predictors, or scale mismatches between the predictors and observations are not addressed in this assessment, but are considered during the data preparation phase prior to application of the framework.

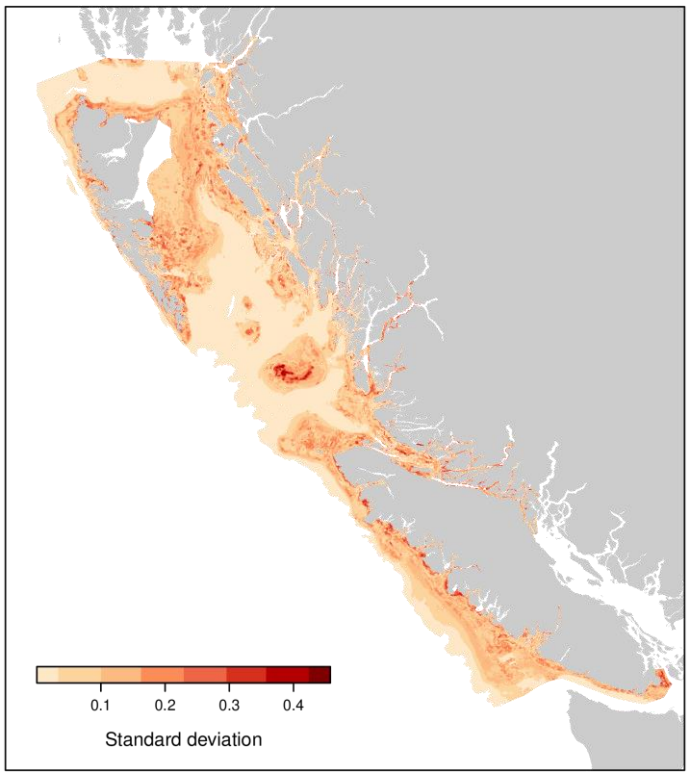

Figure 3.12. Standard deviation of probability of occurrence predictions from five-fold cross-validation models.

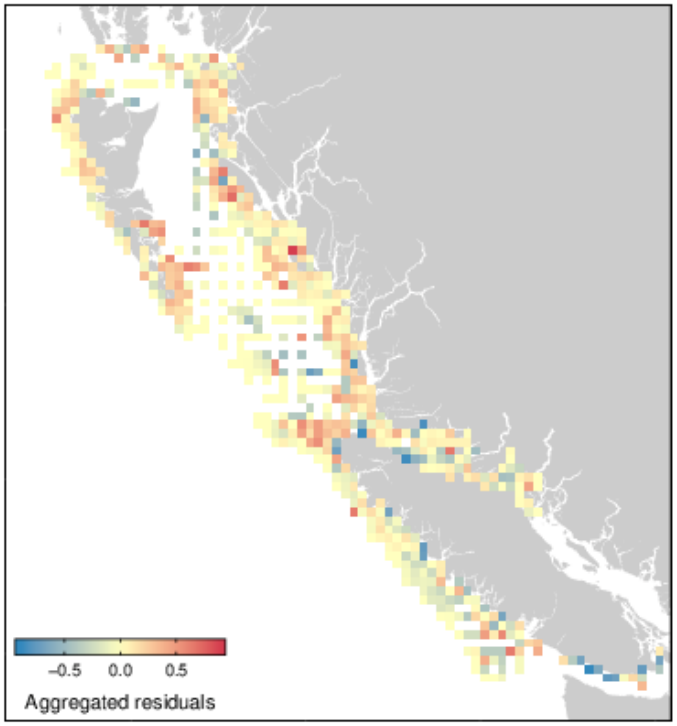

Figure 3.13. Mean model residuals (aggregated to 10 km) from five-fold cross-validation models.

3.7 MODEL INTERPRETATION

Once the models have been fit, hypotheses can be made about the relative strengths and interactions of the many processes that may drive the species' distributions. Such hypotheses are based on examining the relative influence of predictor variables (Figure 3.14) and the marginal effects of each predictor on the observations (Figure 3.15). These graphs can also help formulate a qualitative measure of model performance. For example, unexpected relationships in marginal effects (e.g., a positive slope when a negative slope was expected) signal potential data quality issues (see Section 2 for examples) in the observations or predictors. Additionally, when using spatial blocking, large deviations in either the relative influence or marginal effects across CV models can be indicative of non-stationarity. Importantly, these are only informative measures of stationarity at the scale of model predictions when observations are distributed across the study area.

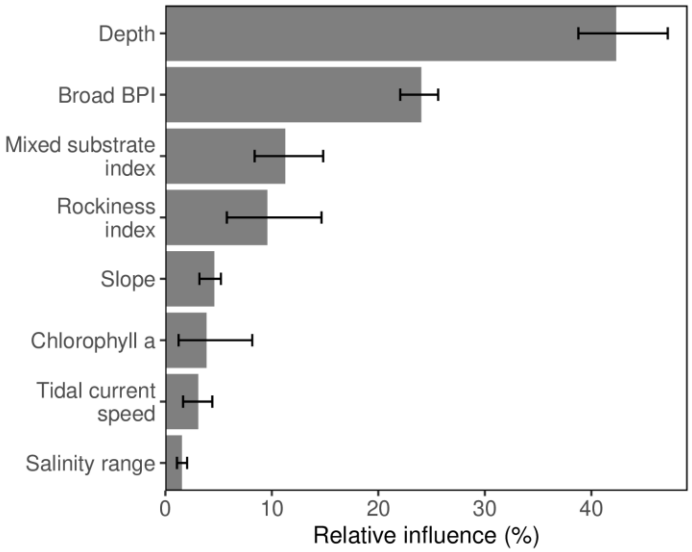

Figure 3.14. Mean relative influence of predictors from five-fold cross-validation models. Error bars represent the minimum and maximum relative influence across the five CV models.

How the relative influence of predictor variables in a model are estimated depends on the modelling method. Regression methods often use standardized coefficients but some have found these unreliable (Bini et al. 2009; Hawkins 2012). An alternative method, used in this framework, is the drop one procedure, where the residual deviance is calculated by dropping each predictor from the model one at a time (Grömping 2015). The increase in residual deviance when a predictor is dropped is an estimate of its relative importance in the model. With machine learning models, approaches for estimating relative influence vary. For example with BRTs, the relative influence of each predictor is estimated by the number of times the predictor splits a branch, weighted by the contribution of each split (Friedman and Meulman 2003).

Marginal effects represent the change in probability of occurrence predictions due to a single predictor. It is estimated by varying a single predictor while keeping the others at their mean values derived from the training data used to fit the models. The framework calculates marginal effects for each predictor using Elith et al.'s (2005) evaluation strip method, a robust and consistent approach to calculating marginal effects suitable to any model fitting method.

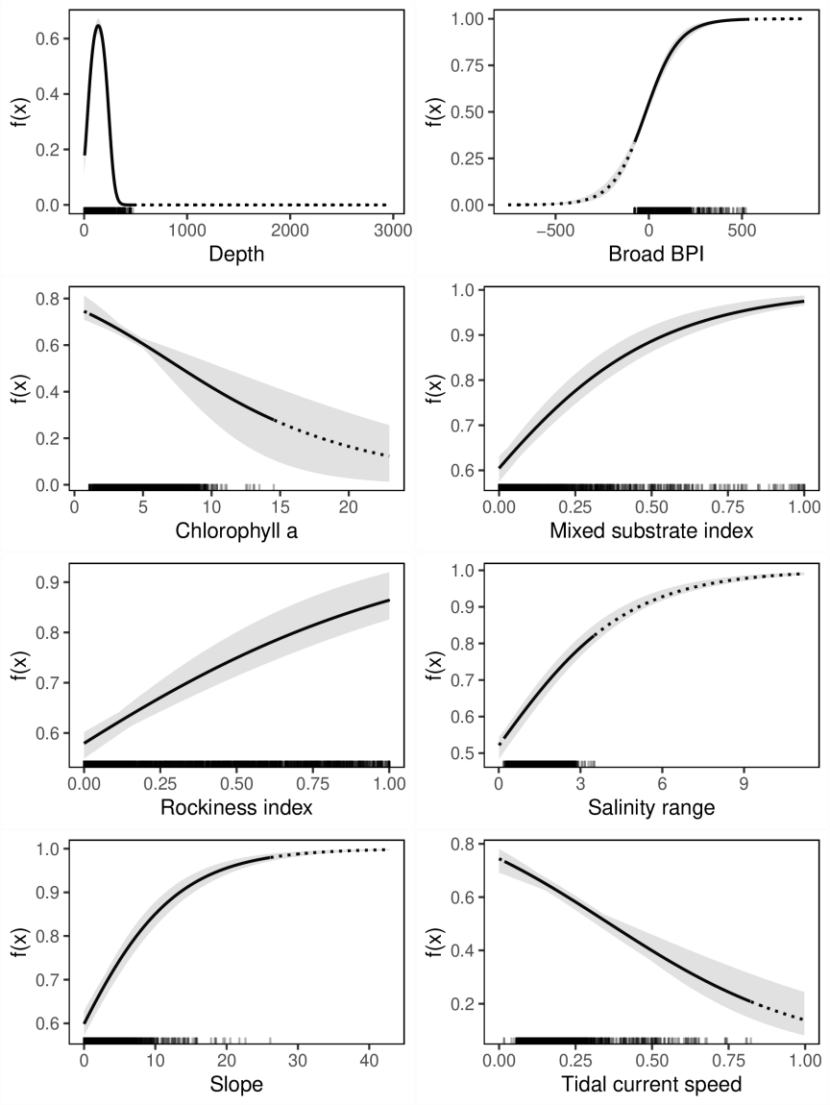

Figure 3.15. Mean, minimum and maximum marginal effects between each predictor variable and the predicted response from five-fold cross-validation models. Marginal effects were calculated from generalized linear models. The black solid line represents the mean, and the grey shaded area represents the minimum and maximum marginal effects across the five models. Marginal effects are presented across the entire range of environmental predictor layers. The dotted line denotes extrapolation in predict space where the predictor variable is outside the range of values used for model fitting indicating there were no species observations in this space. Tick marks along the x-axis indicate the density of observations across the range of predictor values.

3.8 THE ROLE OF SPATIAL STRUCTURE

Understanding the SAC pattern in the observations and model residuals can also contribute to model interpretation. Significant SAC in residuals suggests that the spatial structure in the observations was not fully explained by the predictors (Elith and Leathwick 2009). This may occur when important predictors are missing, when available predictors are poorly scaled to the observations, or when the spatial pattern is influenced by other factors such as biological interactions or barriers to movement (Bahn and McGill 2007; Dormann et al. 2007; Elith and Leathwick 2009; Soberón and Nakamura 2009).

This framework employs variogram plots and the Geary's C statistic to test for positive SAC (Figure 3.16). Variograms illustrate the structure and range of influence of SAC (Legendre and Fortin 1989), where the range of positive SAC is the distance at which semivariance reaches an asymptote. Geary's C is a related metric used to test for the statistical significance of SAC within a defined neighbourhood. The statistic ranges from 0 to greater than 1 with values less than 1 representing positive SAC and values near 1 representing minimal SAC. Geary's C is similar to the often used Moran's I statistic but is less sensitive to extreme values (Legendre and Fortin 1989).

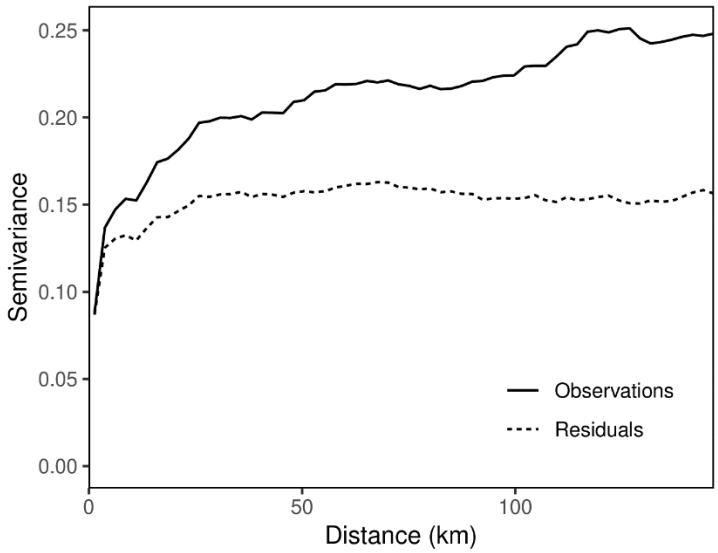

Figure 3.16. Variogram of species presence-absence observations and model residuals.

When model residuals retain spatial pattern, one way to account for the unexplained variation is to include an explicit spatial term in the model (Dormann et al. 2007; Merow et al. 2014). Including a measure of spatial structure can aid model interpretation by providing a better understanding of the relative influence of purely spatial versus environmental predictors (Hothorn et al. 2011; Keitt et al. 2002; Legendre 1993; Lichstein et al. 2002). This framework examines spatial structure in the observations using the auto-covariate method, where a spatial auto-covariate predictor term is derived *a priori* from the observations and included in the model (Hughes et al. 2011). Following Bardos et al. (2015), the auto-covariate term is calculated as the sum of nearby observations within a predefined neighbourhood weighted by symmetrical inverse distance weights (Figure 3.17). In this framework, the neighbourhood is defined as the maximum nearest neighbour distance between observations.

Secondary models built with an auto-covariate term are produced to support model interpretation if significant positive SAC is found in the residuals of the initial models (see workflow in Figure 3.1). Interpreting these auto-covariate models can help improve understanding by illustrating how relationships between observations and environmental predictors as well as the relative importance of predictors can vary when local spatial structure is accounted for (Figure 3.18). This framework does not recommend the use of auto-covariate models for prediction when the sampling distribution of the observations do not support the derivation of a spatially comprehensive auto-covariate predictor.

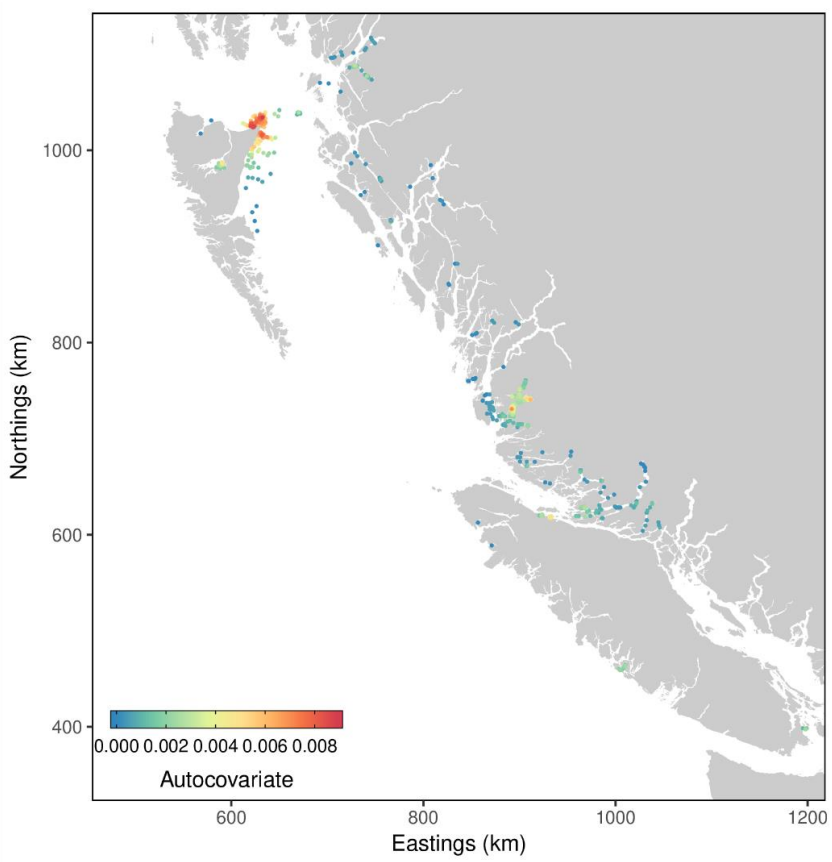

Figure 3.17. Distribution of a spatial auto-covariate predictor that represents the distance weighted sum of nearby observations.

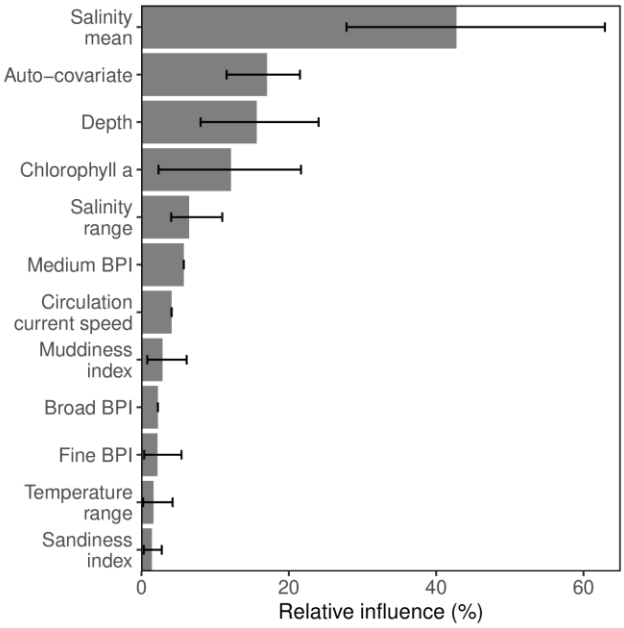

Figure 3.18. Mean relative influence of predictors from five-fold cross-validation models including the auto-covariate. Error bars represent the minimum and maximum relative influence across the five models.

4 AN APPLICATION OF THE SPECIES DISTRIBUTION MODELLING FRAMEWORK

4.1 INTRODUCTION

In 2017, the Canadian Coast Guard began leading the development of a Regional Response Plan for the Northern Shelf Bioregion. This plan is being developed collaboratively with federal, provincial and First Nation partners and is intended to guide emergency response to a marine oil spill incident. In support of this plan, DFO Science has been tasked with filling in data gaps to help identify locations of species of concern.

To address the management need for a well-resolved description of species presence during an oil spill incident, the SDM framework (Section 3) was used to develop habitat models for a diverse set of species in both the Northern and Southern Shelf Bioregions. To maximize management relevance, species were selected from a larger group previously identified as conservation priorities (Gale et al. 2019) or highly vulnerable to oil (Hannah et al. 2017). Benthic species were selected as they are the most susceptible to fouling from oil as it settles. Species were also chosen to represent a diversity of life history characteristics, habitats, and ecological communities, with different levels of data availability and quality, to assess how the framework performed with different combinations of ecological and data models.

This application of the framework uses three model approaches: habitat suitability index (HSI) models; generalized linear models (GLMs); and boosted regression trees (BRTs). These methods span the range of model complexity, and each comes with different strengths and weaknesses. All three methods have been used to model the distribution of aquatic organisms (HSI: Brooks 1997; Raleigh et al. 1986); (GLM: Beger and Possingham 2008; Santoul et al. 2005); (BRT: Leathwick et al. 2006; Leathwick et al. 2008).

HSIs are simple, knowledge-based envelope models that depend on a clear statement of the evidence supporting the processes believed to be responsible for creating suitable habitat (Brooks 1997; USFWS 1981). They draw on published relationships and expert consultation, and are built manually. While they do not require any observational data for development, best practices recommend a calibration process with occurrence data when available (Brooks 1997). In the absence of occurrence data, HSIs can serve both as an ecological baseline, an initial model against which data-driven models can be compared, or can be combined with more complex models in an ensemble (Jones-Farrand et al. 2011).

GLMs represent the class of widely used and well-developed regression methods used to relate species observations to environmental predictors. GLMs require functional forms to be specified, support optimum (using quadratic functions) or threshold values, and can include interactions in an intuitive manner. GLMs were chosen over the more complex generalized additive models as these models can approach the complexity of machine learning methods, and a goal here was to allow models of different complexities to be compared.

Boosted Regression Trees (BRTs) are a powerful tree-based machine learning method for SDM development. Regression trees recursively divide the response data into homogenous groups, with each tree split based on a single predictor variable (De'Ath and Fabricius 2000). Boosting is a stepwise process that builds several trees at each step to model the residuals of the previous iteration (Elith et al. 2008). Each step uses a random fraction of the training data (bag fraction) to build the new trees, improving model performance at each step (Elith et al. 2008). Interactions are represented implicitly, based on the structure of the branches in the trees. However, this can make them more difficult to interpret. BRTs have a variety of settings that influence the complexity of the final model (e.g., tree complexity, learning rate, number of starting trees, maximum trees to build and stopping condition).

Each of these methods represents a different approach to SDM development, and allows different types of inferences to be drawn. This application of the SDM framework, using different ecological, data, and statistical model approaches, illustrates how the framework performs in different model contexts, providing guidance for the development of other SDMs in Canadian waters. The species models developed in this application of the SDM framework are intended for oil spill response purposes and would require further review (see Figure 1.1) for use in other management applications.

4.2 METHODS

This section describes the application of each of the steps in the framework (i.e., data preparation, model fitting, evaluation, and prediction; Figure 1.1) to the species of interest. Methodological details are included on the study area, species modelled, occurrence and predictor data used, and modelling methods applied.

4.2.1 Study areas

The framework was applied to the shelf waters of Canada's Pacific Coast at two resolutions (Figure 4.1). The shelf study area (modelled at a spatial resolution of 100 m) is bounded by 47°57' - 55°58'N and 123°6' - 134°19'W. It covers the Northern Shelf and Southern Shelf marine Bioregions (DFO 2009), with the western boundary following the base of the shelf slope (roughly the 2000 m isobath). The Strait of Georgia marine bioregion was excluded because the shelf environmental predictors used in this analysis do not adequately resolve the oceanographic features in this bioregion, and higher resolution predictors have been built specifically for that area. Future work will apply this framework to the Strait of Georgia using these higher resolution data.

The nearshore North Central Coast study area (modelled at a spatial resolution of 20 m) is located within the Northern Shelf Bioregion and is restricted to the nearshore subtidal region from the high water line to 50 m depth, and within 5 km from shore (Gregr et al. 2013). The nearshore area extends from Portland Canal in the north, south to Queen Charlotte Strait, and covers over 17,500 km of coastline with many islands, channels, and fjords. This area was selected as the pilot study area for this work because of a higher density of species occurrence data collected as part of earlier proposed industrial developments in the region.

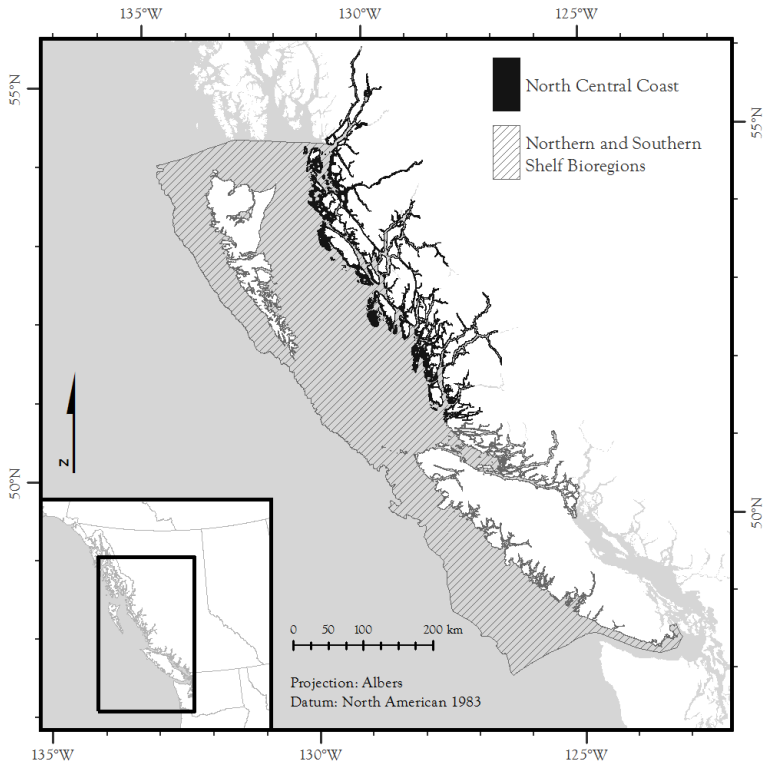

Figure 4.1. The two study areas used in this analysis. The nearshore (20 m resolution) North Central Coast study area covers approximately 5,620 km^2. The shelf study area (100 m resolution) includes the Northern Shelf and the Southern Shelf marine bioregions and covers approximately 127,840 km^2.

4.2.2 Environmental predictors

Bathymetry and derivatives

A total of six layers were derived from various bathymetric data sources (Carignan et al. 2013; Davies et al. 2019; Gregr 2012) with different spatial resolutions (Table 4.1). Using ArcGIS 10.4, the bathymetry layers were resampled and mosaicked to produce depth rasters for each of the two study areas. All environmental predictors were projected in the B.C. Albers projection (EPSG:3005).

Slope, rugosity, and bathymetric position index (BPI) variables were derived from the depth rasters at 20 m and 100 m spatial resolutions. Slope (the maximum rate of change in depth across each cell) was calculated in degrees. Rugosity (an index of surface roughness used as a measure of structural complexity) was calculated according to Du Preez (2015). Various methods exist to calculate complexity. We chose the arc-chord ratio (ACR) rugosity method (Du Preez 2015) because it decouples rugosity from slope, making it a superior measure of surface roughness (Du Preez et al. 2016). BPI values represent a cell's topographic position relative to a neighbourhood of cells and may be positive (e.g., ridges or crests) or negative (e.g., valley bottoms). To capture local and broader-scale topographic features within the data, BPI layers were created at fine, medium, and broad scales using varying neighbourhoods (Walbridge et al. 2018). Nearshore fine BPI was calculated with a neighbourhood distance from 60-500 m, and the shelf used a range from 300 - 2,500 m. The neighbourhoods for the nearshore and shelf medium BPI layers were 200 - 2,000 m and 1,000 - 10,500 m respectively. The neighbourhoods for the nearshore and shelf broad BPI layers were 500 - 5,000 m and 2,500 - 25,000 m respectively.

Oceanographic layers

A total of six layers describing bottom temperature (in degrees Celsius), salinity (in practical salinity unit, or PSU), and current speeds (tidal and circulation in metres per second) were derived from a Regional Ocean Modelling System (ROMS) circulation model of Canada's Pacific coast (Table 4.1, Masson and Fine 2012). The ROMS model has a spatial resolution of approximately three km and its domain extends from the Columbia River to southeastern Alaska. ROMS model output variables were sourced from a hindcast for the period of 1998 – 2007. The bottom layer was represented by the deepest of 30 sigma levels, which followed bottom depth.

Current speed was calculated from mean zonal (u) and meridional (v) velocities using a root mean square method after the velocities were spatially aligned by shifting them horizontally with linear interpolation. Temperature, salinity, and circulation current speed were sourced from 15 day means while average monthly tidal current speed was calculated from three hour means. The finer three-hour temporal resolution of the component velocities allowed tidal current to be resolved, while the coarser 15-day means effectively resolve non-tidal ocean circulation current.

Temperature, salinity, and current speeds (tidal and circulation) were then averaged temporally for the spring/summer season, from April to September, over the ten-year time period. The spring/summer season was chosen to align with when the majority of species observations were collected. Minimum and maximum values were calculated over the ten years to represent the temporal range for temperature and salinity.

Chlorophyll

Mean surface chlorophyll-a data, a proxy for primary productivity, was derived from the one km resolution MODIS L2 product. NASA-derived chlorophyll a concentration (mg/m^3) were calculated from reflectance using the OC4 and CI algorithms (for more details see Hu et al. 2012). Daily swath data between March and October from 2012 to 2015 were downloaded and mosaicked by month. Months from November to February were excluded as cloud cover is persistent during that period. The monthly data were interpolated spatially using Spline with Barriers (ESRI 2019) with the coastal high water line as a barrier to fill in any gaps that remained after mosaicking. These gaps were typically located nearshore and in coastal inlets. Mean chlorophyll-a concentration (mg/m^3) was calculated by averaging the interpolated monthly mosaics across the four-year period.

Table 4.1. Sources of environmental predictor variables used to model habitat suitability for 12 species at 2 spatial resolutions (20 m for the nearshore and 100 m for the shelf). See section 4.2.2 for a description of the units for each predictor layer.

Environmental data type	Predictor layer(s)	Source(s)	Native resolution(s)	Years	Study area(s)	Layers (N)
Bathymetry	Bathymetry Slope Rugosity Broad BPI Medium BPI Fine BPI	British Columbia 3 arc-second Bathymetric DEM (Carignan et al. 2013) 100 m DEM (Gregr 2012) Bathymetric elevation models (Davies et al. 2019)	3 arc-seconds 100 m 20 m	1930-2012	shelf shelf nearshore	6
Oceanographic	Mean summer bottom salinity Bottom salinity range Mean summer bottom temperature Bottom temperature range Mean summer tidal speed Mean summer circulation	Regional circulation model of BC (Masson and Fine 2012)	3 km	1998–2007	nearshore shelf	6
Chlorophyll-a	Mean	NASA Ocean Color	1 km	2012-2015	shelf	1
Fetch	Sum fetch Minimum fetch	Python script (Gregr 2014)	50 m		nearshore	2
Substrate	Rocky Mixed Sandy Muddy	Background Substrate (Gregr and Haggarty 2017)	20 m, 100 m		nearshore shelf	4

Fetch

Fetch is the distance over which wind-driven waves can build and provides a proxy measure of exposure to wave action (Burrows 2012). Westerbrom and Jattu (2006) claim that wave exposure is an important factor for determining the distribution of marine species along rocky shores, and Lessard and Campbell (2007) demonstrated the importance of fetch to abalone distributions. It was therefore critical to include a measure of exposure in the nearshore models.

Fetch values were calculated for 1,171,638 points (one every 50 m grid cell) covering the nearshore North Central Coast study area using a custom Python script (Gregr 2014). For each point, 72 bearing lines were generated and extended to a maximum of 200 km at every 5 degree bearing. Fetch lines were then clipped by land using the coastal high water line. Sum of fetch (a proxy for wind-wave exposure) was then defined as the cumulative fetch line length (in m) for each point.

The sum of fetch values were converted to rasters with 50 m cell size. To prepare the final raster layer, the 50 m layer was resampled to 20 m using the bilinear method and aligned with the other predictor layers. A similar procedure was used for calculating minimum fetch (a proxy for distance to land), using the smallest fetch value (in m) for each point.

Categorical substrate and neighbourhood substrate layers

Following on the work of Gregr and Haggarty (2017), categorical substrate models were built for both study areas. The models predict four categories: rock, mixed, sand, and mud. The models were built using the Random Forest approach, a classification method that uses training data and predictor layers (Breiman 2001). Substrate observation data were obtained from various sources and coded into the substrate categories. To prepare the observations for modelling, they were randomly split (while preserving prevalence) into training and testing data. Bathymetry and its derivatives, along with fetch, and ocean circulation were used as predictors. Model fitting and prediction were carried out using the training data, while the withheld testing data (one third of the observations) were used to evaluate the model and generate performance statistics. Code for model fitting, prediction, and evaluation was written in the R statistical language (R Core Team 2018) and used the randomForest package (Liaw and Wiener 2002).

We derived continuous neighbourhood substrate predictors from the categorical substrate models to reflect uncertainty in the categorical models, and better characterise substrate in the general locale rather than in a single cell - an attribute that may be more relevant (i.e., proximate) to more mobile target species. Rockfish habitat, for example, may be more effectively represented using a local rockiness index rather than a categorical layer. Continuous, neighbourhood index layers were created for each category (rocky, mixed, sandy, muddy) at both spatial resolutions (20 m and 100 m) using the following steps. First, a binary layer was created for each of the substrate categories (rock, mixed, sand, mud). For each of the binary layers, the focal sum was then calculated using a circular neighbourhood with a 10-cell radius. Finally, the focal sum layers were converted to index layers by rescaling them onto [0, 1]. The resulting index layers provide a measure of substrate density within the neighbourhood of cells.

Spatial predictors

All models presented in this application were produced with environmental predictors only. Spatial predictors layers (e.g., auto-covariates) discussed in the framework were not explored in this application.

4.2.3 Species observations

Data selection

Twelve benthic species were selected for this application. Eight species were modelled in the nearshore study area (20 m resolution): Northern Abalone (*Haliotis kamtschatkana*), Pacific Geoduck (*Panopea generosa*), Red Sea Urchin (*Mesocentrotus franciscanus*), Eelgrass (*Zostera* spp.), Pterygophora Kelp (*Pterygophora californica*), Ochre Sea Star (*Pisaster ochraceus*), Blue Mussel complex (*Mytilus edulis*, *M. trossulus*, and *M. galloprovincialis*) and Littleneck Clam (*Leukoma staminea*) and four species were modelled in the shelf study area (100 m resolution): Dungeness Crab (*Metacarcinus magister*), Quillback Rockfish (*Sebastes maliger*), Yelloweye Rockfish (*Sebastes ruberrimus*) and Orange Sea Pen (*Ptilosarcus gurneyi*).

When selecting species occurrence data, sample size, extent, sampling bias and precision were considered as discussed in Section 2.1. The same general approach was applied to each species. First DFO shellfish and groundfish data holdings were searched for targeted or systematic synoptic surveys that adequately sampled the species of interest. For a species to be deemed adequately sampled, the survey method and gear type were required to sample the species so that the assumption of absence given a lack of presence holds true. This determination was based on species expert's knowledge of the detectability of the species given the survey gear, and the prevalence of the species in the survey observations. Once a survey was deemed appropriate, sample size was evaluated to determine whether additional data were needed.

Additional survey data were deemed necessary when sample sizes for either presence or absence observations were low (e.g., < 100 absence observations) or when the spatial coverage of observations across the study area was limited in either geographic (e.g., no observations in the southern portion) or predictor space (e.g., all observations occurred on rocky habitat). Although best practice is to use a single survey to source species occurrence data, in many cases a single survey alone is insufficient to represent species occurrences across the desired study area (Table 4.2). When supplementing species occurrence data with additional surveys, attempts were made to find surveys with similar detectability rates. For all species, occurrence data were sourced from surveys using the same gear type (e.g., dive surveys for nearshore species, longlines for Rockfish species and baited trap for Dungeness Crab). While the goal was to maximize coverage across the study area and habitat types, not all species were sampled comprehensively across their range and the study area (e.g., Dungeness Crab observations were mainly restricted to muddy, sandy nearshore and shallow areas, see Figure A.13).

Once a species occurrence dataset was determined to be sufficient, additional data were not included, even if they were available. For example, Rockfish trawl data were available in addition to the longline survey data. However, including the trawl data would introduce bias stemming from differences in catchability and area sampled. Such differences in gear types (e.g., trap vs. trawl) influence the detectability of species (e.g., Wells et al. 2008), effectively leading to an observation bias. While the significance of this effect could be tested for *a priori*, and may give an estimate of gear type effect, the result would be confounded because of its correlation with habitat (e.g., trawls are used over deeper, softer sediments while longlines are used over shallower, hard bottoms). Disentangling the gear effect from the habitat effect would require a separate analysis and significant spatial overlap between the surveys, an unlikely situation given the underlying reasons (i.e., gear loss) for the use of different gear in the first place. Further, predicting to new areas with a species distribution model that included a gear type predictor would require an associated (and unavailable) layer describing the distribution of

gear types or alternatively require the modeller to choose which gear type would be more appropriate for prediction.

Four of the twelve species selected (Ochre Sea Star, Blue Mussel, Littleneck Clam and Orange Sea Pen) did not have adequate occurrence data for modelling within our study areas. While data were available for all four species, they were determined inappropriate because of low sample sizes (e.g., < 100 presence observations), low precision in the spatial location of the observations (e.g., site versus quadrat) or low taxonomic resolution (e.g., observations at the order level).

Spatial data preparation

The spatial precision of species observations varied by survey (e.g., only the start location of sampling versus the start and end location). When deciding how best to represent such observations spatially, we considered the precision of the spatial location of the observations relative to the resolution of predictor data. Thus, to make the best use of the 20 m resolution predictor data, the nearshore species data needed higher precision in spatial information than the shelf species data.

All nearshore species occurrence data were sourced from SCUBA dive surveys (Table 4.2). All dive surveys followed a transect-based protocol with transects perpendicular to the shoreline, with the exception of the Abalone survey. Observations were recorded for quadrats placed along a transect, running from deep to shallow, however only transect start and end points were recorded. To represent nearshore species at the 20 m resolution, quadrat positions along the transects were needed. To estimate the position of quadrat observations, points were created every 20 m along the transect, and depth values were extracted from the bathymetry layer for each point. Quadrat observations were then assigned to the points that most closely matched the observed (corrected for tide) quadrat depth. If multiple quadrats were assigned to the same point, the observations were aggregated. If the difference between recorded quadrat depth and the 20 m bathymetry was greater than 10 m, the observation was removed from the dataset.

DFO dive surveys for Abalone follow a grid-based survey protocol (Breen and Adkins 1979) rather than a transect-based protocol. Observations are recorded per quadrat, with quadrats placed in a 4 x 4 grid pattern (covering an area of 16 by 7 m) in habitat considered to be suitable for Abalone. The start position of the sampling is recorded. For this survey, individual quadrats were aggregated by site and the starting position was used to represent their spatial location. Presence-absence observations were aggregated by assigning presence when at least one presence observation occurred within a quadrat.

Species occurrence data for the shelf study area were sourced from surveys using longline and baited trap sampling methods (Table 4.2). For longline fishing events, spatial locations were represented by line segments between the start and end position of the fishing event. For Dungeness Crab surveys, spatial locations of the baited trap lines were represented by the start position as this was the best available spatial information.

Table 4.2. Summary of the occurrence data selected to model the twelve species in this study. Generalized linear regression and boosted regression trees were used to build models for the eight species not found data deficient (DD). Surveys listed under data sources (ABL = Abalone; RSU = Red Sea Urchin; BHM = benthic habitat mapping; Cuke = Sea Cucumber; HBLL=hard bottom long line; IPHC=International Pacific Halibut Commission) represent scientific surveys for stock assessment or monitoring lead by DFO Pacific and industry partners.

Species	Nearshore or Shelf (N/S)	Sample size	Prevalence (%)	Years	Data sources (presence-absence)	Data sources (absence only)	Spatial data type	Spatial data precision
Northern Abalone (*Haliotis kamtschatkana*)	N	2,293	22	2011-2016	ABL, RSU, BHM	-	Points	Site location (ABL) or estimated location along transect (RSU, BHM)
Pacific Geoduck (*Panopea generosa*)	N	9,350	58	2010-2017	GDK, BHM	RSU, Cuke	Points	Estimated location along transect
Pterygophora Kelp (*Pterygophora californica*)	N	6,607	3	2010-2017	ABL, Cuke, RSU, BHM	-	Points	Estimated location along transect
Red Sea Urchin (*Mesocentrotus franciscanus*)	N	3,300	26	2010-2016	RSU, BHM	GDK, Cuke	Points	Estimated location along transect
Eelgrass (*Zostera* spp.)	N	12,567	4	2010-2017	GDK, Cuke, RSU, BHM	-	Points	Estimated location along transect
Dungeness Crab (*Metacarcinus magister*)	S	391	49	1982-2009	Dungeness Crab surveys	-	Points	Start position of gear deployment
Quillback Rockfish (*Sebastes maliger*)	S	4,937	41	2003-2018	HBLL, IPHC	-	Lines	Start and end position of longline gear
Yelloweye Rockfish (*Sebastes ruberrimus*)	S	4,937	51	2003-2018	HBLL, IPHC	-	Lines	Start and end position of longline gear
Orange Sea Pen (*Ptilosarcus gurneyi*)	S	DD	-	-	-	-	-	-
Pacific Littleneck Clam (*Leukoma staminea*)	N	DD	-	-	-	-	-	-
Ochre Sea Star (*Pisaster ochraceus*)	N	DD	-	-	-	-	-	-
Blue Mussel complex (*Mytilus edulis, M. trossulus, M. galloprovincialis*)	N	DD	-	-	-	-	-	-

Dive survey assumptions

Several assumptions were made when selecting species occurrence data sourced from dive surveys. In many dive survey protocols only dominant algal species are recorded in each quadrat, potentially creating false absences. Expert consultation indicated that when both Kelp and Eelgrass are present, they are likely to be dominant. Therefore we made the assumption that the lack of a Kelp or Eelgrass presence represented a true absence of that species.

Non-target species in some dive survey protocols are only recorded at the site or transect level, not the quadrat level. However, we required more precise observations (at the quadrat level) for our nearshore models. Consequently, absences at the transect level were assumed to indicate an absence for that species in all quadrats, but presences at the transect level were not used to indicate presence at all quadrats.

4.2.4 Data preparation

The preparation of the species observations and predictor data for modelling were completed following the methods outlined in Section 3.1. Here we highlight additional details that are specific to this application.

Mean summer temperature and salinity were correlated, falling slightly outside the variance inflation factor (VIF) threshold of ten. However, both temperature and salinity were retained for consistency across species models. Not all substrate predictors were included for every species because the sandy and rocky predictors were highly inversely correlated. Thus, for species known to inhabit rocky areas, the rockiness index was used, while for species known to inhabit soft sediments, the sandy and muddy predictors were used. The mixed substrate predictor was used in all species models.

Five-fold spatial block CV was used to build GLM and BRT models and validate all models, except those for species with insufficient occurrence data. For each species, identical CV folds were used for training and testing of these models. Using identical testing data allows a direct comparison of model validation metrics from models build with different methods. The size of spatial blocks was based on the median range of SAC across all the environmental predictor layers. Shelf spatial blocks were approximately 75 km by 75 km and nearshore spatial blocks were 35 km by 35 km.

4.2.5 Habitat suitability index models

HSI models were built for each of the twelve species based on literature review and consultation with species experts. In contrast to the correlative models (GLM and BRT), HSI models did not use species observations (described in Section 4.2.3) to develop relationships with environmental predictors. Rather, for each relevant environmental predictor, its contribution to habitat suitability was selected from a set of common relationships (e.g., Figure 3.4). Each relationship was then applied to the corresponding environmental predictor producing a univariate habitat suitability prediction ranging from 0 to 1 (unsuitable to fully suitable). Predictor relationships were developed for the environmental predictor layers described in Section 4.2.2. For species (i.e., Red Sea Urchin) known to shift habitat to deeper waters in areas of higher exposure (due to wind-wave action), this interaction was represented by dividing coastal habitats into low and high exposure areas using a cut-off of 20 (100s of km), and defining slightly different depth preference curves for each of the two areas (see Appendix B for details).

The collection of univariate predicted suitabilities were then combined using the limiting factor method (USFWS 1981). This method uses the predictor with the lowest suitability as the overall suitability for each cell. Unlike compensatory relationships, high suitability in one predictor

cannot compensate for the low suitability in others. To calculate the relative influence of each environmental predictor in HSI models, the number of cells where each predictor was limiting was divided by the total number of cells in the study area.

Building the HSI models was an iterative process, where preliminary models were built based on ecological relationships in the literature and then refined through expert consultation. Documents for expert consultation were prepared, describing the preliminary HSI models and the ecological rationale behind them, to facilitate the consultation process and maintain consistency in the advice being sought. There were varying levels of information available for each species. For some species, experts drew on their general understanding of the species distribution and its relationships with environmental predictors. For other species, more specific information was available to the experts, who refined environmental predictor relationships using environmental ranges from survey data and verified final model outputs using known areas of occurrence or density estimates. As such, the role of species observations in the HSI models varied, with some models being entirely independent of species observations (e.g., Abalone and Eelgrass), while in others, species observations supported the development of the univariate relationships (e.g., Quillback Rockfish and Geoduck).

The HSI model development process differed from that of the data-driven models (GLM and BRT) in that only one HSI model was built for each species. Consequently, no estimate of variation across model predictions could be made. Once the models were finalized, eight of the twelve HSI models were evaluated with testing data from five-fold spatial block CV, allowing means and standard deviations of model validation metrics to be calculated. There were no suitable data available for evaluating the remaining four models (Table 4.2).

4.2.6 Generalized linear models

Five GLMs, one for each CV run, were built for each species. A binomial error distribution and logit link function were used as is recommended for binary observations (Guisan et al. 2017; Tabachnick et al. 2007). Selecting the best model for each CV dataset was automated with the dredge function (Barton 2018). Dredge performs model selection by fitting every possible combination of predictors given a global model and ranking the models based on an information criterion. In this case Bayesian information criterion (BIC) was used. The global model included every predictor from the predictor dataset as well as specific quadratic or interaction terms that were believed to be ecologically important for a specific species, as determined by the expert-derived information from HSI models. For example, Northern Abalone are known to exhibit a quadratic relationship with exposure where habitat suitability peaks at medium exposure levels (Lessard and Campbell 2007). Quadratic and interaction terms were restricted in this way because of computational limitations to dredging with a full suite of higher order and interaction terms. The best models from each of the five CV runs, each with a potentially different model structure, were evaluated using the testing fold, and then used to predict the probability of occurrence within the study area. The mean and standard deviation of evaluation metrics and prediction surfaces were calculated across the five models.

4.2.7 Boosted regression trees

Five BRT models, one for each CV run, were built for each species. BRT models were built with functions from gbm (Greenwell et al. 2018) and dismo (Hijmans et al. 2017) R packages, using a Bernoulli distribution for the loss function (Ridgeway 2007). BRT models require a number of initial parameters that control how the model is structured (Table 4.3). Parameter values were based on Elith et al. (2008), except that the bag fraction was not tuned because initial tests showed that varying the bag fraction did not affect the BRT performance.

Table 4.3. Initial parameters for boosted regression tree models.

Parameter description	Initial value
Maximum number of trees to fit	20,000
Number of initial trees	50
Number of trees per step	50
Optimal number of trees	Calculated during tuning
Number of cross-validation folds for tuning	5
Tolerance for stopping condition during model tuning	0.001
Learning rate (contribution of each tree)	0.05, 0.01, 0.005, 0.001
Tree complexity (number of splits or nodes)	2,4,6,8
Bag fraction (proportion of training data selected at each step)	0.5

The total number of trees to build, the learning rate and tree composition all required tuning to determine their optimal values. The tuning procedure fit 16 preliminary models, one for each unique combination of learning rate and tree complexity values (Table 4.3), for each CV run. Within the model tuning procedure the training data was split using five-fold random CV into internal training and holdout data to evaluate the stopping criterion that determines when the optimal number of trees has been reached. Excluding a portion of the training data to use for the stopping condition, as opposed to using the testing fold, keeps the testing data independent of model fitting. The tuning procedure built models iteratively, starting with 50 trees and increasing by 50 trees with each iteration until the 20,000 tree maximum or the stopping condition was reached. The stopping criterion, set by the tolerance parameter, controls the allowable relative increase in holdout deviance from the previous iteration with 50 fewer trees to the most recent iteration.

To select the best model from the 16 tuning models, predictive performance was estimated using mean percent deviance explained from the holdout data. The best model and thus the optimal model parameters were determined for each five-fold spatial block CV run. Following the rule of thumb recommended by Elith et al.(2008), models with fewer than 1,000 trees were not considered. Subsequently, five final models were built, one for each of the training CV datasets using the optimal parameters tuned for each dataset. Models were evaluated with testing fold, then used to predict the probability of occurrence within the study area. The mean and standard deviation of the evaluation metrics and prediction surfaces were calculated across the five models.

4.2.8 Ensemble models

Ensemble model predictions were created by averaging the GLM and BRT model predictions, weighted by their performance (see Section 3.5). Five ensemble models, one for each CV run, were created for each species. Creating an ensemble model for each CV run was necessary to ensure accurate model validation: if only a single ensemble model was calculated from the mean predictions of GLM and BRT models, no testing data would remain for evaluating the ensembles. This method ensured that the ensemble models could be evaluated with the same testing data as HSI, GLM and BRT models. As with the GLM and BRT models, the mean and standard deviation of the evaluation metrics and prediction surfaces were calculated across the five ensemble models.

4.2.9 Model evaluation

Models were evaluated using the metrics presented in Section 3.4. Threshold-dependent statistics were based on thresholds derived by maximizing the TSS using the testing data.

4.3 RESULTS AND DISCUSSION

4.3.1 Model performance by species

The best performing models were for Quillback Rockfish (ensemble AUC = 0.91) and Abalone (ensemble AUC = 0.85) (Figure 4.2). Geoduck, Kelp, Urchin and Eelgrass models all preformed moderately well with ensemble model AUC values greater than or equal to 0.8. The poorest performing models were for Dungeness Crab and Yelloweye Rockfish, although both had adequate ensemble model AUC values between 0.71 and 0.78, respectively. Collectively, the HSI models performed poorly with the majority yielding AUC values less than 0.7. The poorest performing HSI model was Eelgrass with an AUC value less than 0.5 indicating the model was no better than random.

These differences in performance are not solely driven by sample size since both rockfish species have identical sample sizes but different performance metrics. While small sample size likely contributed to the poor performance of the Dungeness Crab model, a more significant factor may have been the limited distribution of observations across geographic and predictor space (Figure A.13). Differences in performance may also be related to aspects of the underlying ecological models, in that stronger and more direct relationships between the occurrence data and a few relatively important predictors tends to lead to higher performing models. For example, the best performing HSI models (Abalone and Quillback Rockfish) both have a strong relationship with a single predictor (see Figure 4.3 and Figure A.18, respectively). These predictors also meet the important criteria of extending past the range of suitable values for these species. In contrast, while the Kelp HSI is also strongly reliant on a single predictor (depth), this relationship is more distal (as depth is likely a proxy for light availability).

Comparing the marginal effects and relative influence plots across the different model methods provides insights into why the HSI models for Eelgrass and Crab performed so poorly (i.e., Figure A.12 and Figure A.15). For Eelgrass, depth and slope were important in the data-driven models (GLM and BRT), while depth and substrate were the main contributors to the HSI (Figure A.12). Although depth was important in both the HSI and data-driven models, the structure of the relationship between Eelgrass and depth differed. Additionally, slope was notably absent from the HSI; it was not included because no related expert guidance was available. This highlights a potential shortcoming of HSI: there can be limited or no knowledge from literature or experts to articulate a functional relationship. For Dungeness Crab, depth and salinity were important drivers in both the HSI and data-driven models (Figure A.15) and, similar to the Eelgrass example, these predictors differed in their functional relationships between the two models. Specifically, the data-driven models showed a decreasing probability of occurrence at high salinities (> 30) and deeper depths (> 100), whereas the HSI identified salinities and depths past those values as highly suitable. These discrepancies may be due to a difference in the scale at which the ecological understanding and correlations are developed. For Dungeness Crab, ecological understanding comes from local, fine resolution observations, at the low end of the salinity range, while the salinity predictors are comparatively coarse spatially (3 km resolution) and temporally (10 year means). While both scales may represent relevant processes, the coarse resolution predictors cannot capture the fine scale ecological understanding, leading to a scaling mismatch as described by Wiens (1989).

These species-specific examples illustrate the diagnostic utility of the performance metrics presented in this framework. A lack of coherence in the relative influence and marginal effects between knowledge-based (e.g., HSI) and data-driven (e.g., GLM) models provides information on potential uncertainties in model structure and the appropriate resolution (either spatial or temporal) of the environmental predictors.

4.3.2 Comparisons of model predictions across methods

As illustrated above, model confidence is increased by consistency in the relative influence of predictors and the shapes of their marginal effects across different methods (e.g., Figure 4.3; and see Appendix A). Species exhibiting such consistency are likely to be well represented (i.e., have higher realism, sensu Araújo et al. 2019) by the predictors. A notable difference between the marginal effects obtained from the GLM and BRT models, despite their similar shapes, is the larger error bounds on the BRT curves (Figure 4.3), representing a larger difference in marginal effects across CV models. This pattern is consistent with our understanding that BRT models fit closer to the training data than GLMs and thus will have higher variation in marginal effects for models built with different training data sets.

Comparisons across models can also help assess the performance of the different methods. The BRT models fit the training data more closely than GLM models for all species (Figure 4.2), but, the two models performed similarly when evaluated with testing data, although for different reasons. The GLMs describe more general patterns (less flexible relationships) so they miss finer scale patterns which BRTs can resolve. In contrast, the BRTs, because they are so flexible, fit too close to the training data, thereby generating patterns less relevant to the testing data. Both methods thus fail to explain some variation in the testing data.

This comparison of the performance of the less flexible GLMs to the more complex BRTs confirms observations in the literature that more complex models (e.g., those fit with machine learning methods) tend to overfit to the training data. However, it also shows that when evaluated using spatially independent data (as obtained through spatial blocking), the measured performance of the BRT models are in line with the less complex GLMs. This both emphasizes the importance of correct model evaluation, and implies that model complexity, in and of itself, is not a barrier to model transferability. This agrees with the literature (e.g., Evans et al. 2013; García-Callejas and Araújo 2016) that suggests over-fitting to biased data can be a bigger problem for model transferability than model complexity.

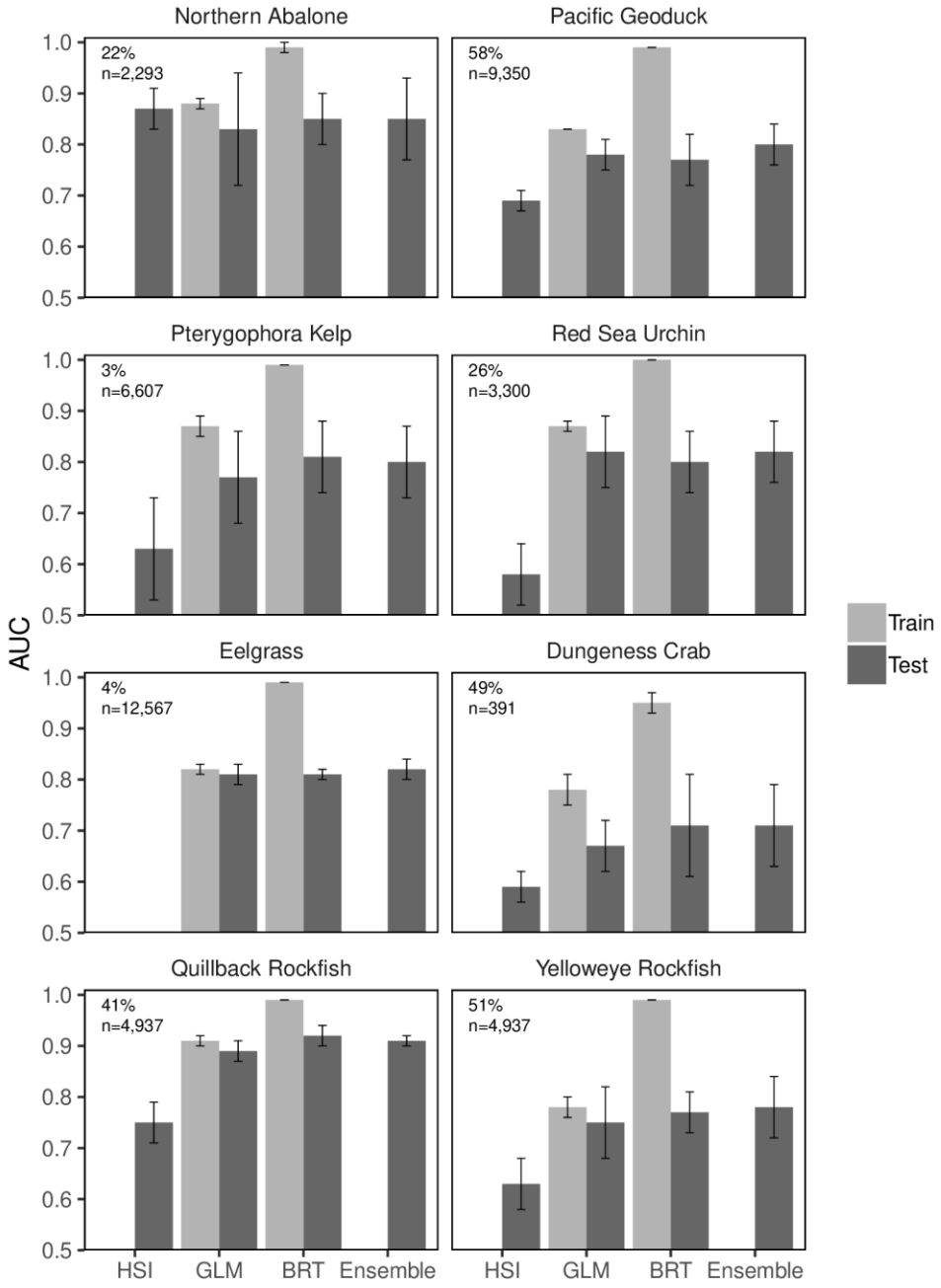

Figure 4.2. Performance of the eight distribution models for species with sufficient occurrence data. Mean area under the receiver operator characteristic curve (AUC) is based on five-fold spatial cross-block validation tests and are shown for each of the four model types: habitat suitability index (HSI), generalized linear model (GLM), boosted regression tree (BRT), and ensemble. Error bars represent one standard deviation. Prevalence and sample size are reported in the upper left corner. Bars not shown have AUC values less than 0.5. HSI and ensemble models were not evaluated with training data.

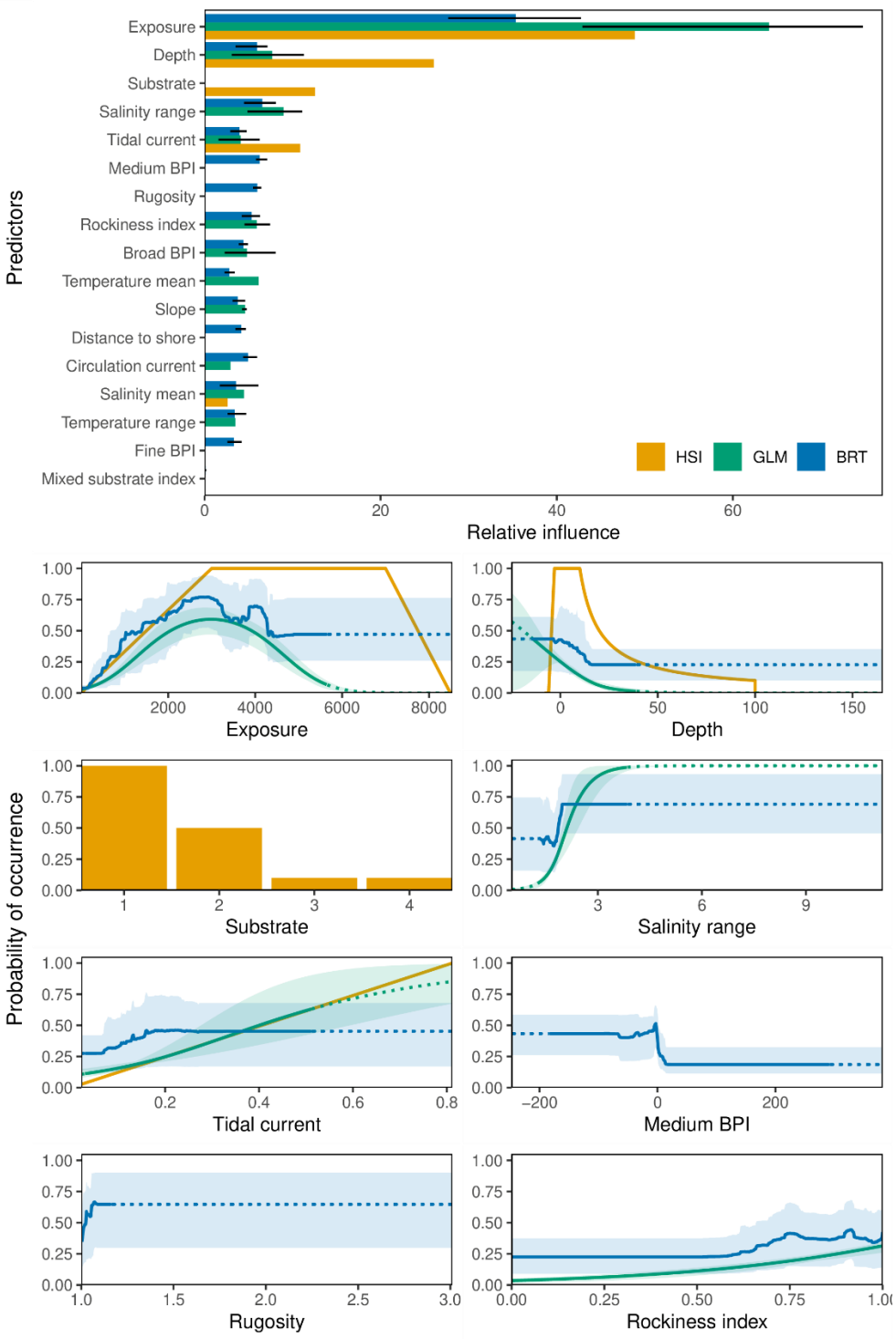

Figure 4.3. Relative influence of predictors and marginal effects of the eight most influential environmental predictors of Abalone distribution from the HSI, GLM and BRT models. For GLM and BRT models, the bars in the relative influence plots represent the mean and the error lines show the minimum and maximum relative influence across the five-fold CV models. In the marginal effects plots, black solid lines represent the mean and the grey shaded area represents the minimum and maximum marginal effects across the five-fold CV models. To represent substrate a categorical substrate layer was used for the HSI model and a continuous rockiness index was used for GLM and BRT models.

In all but three cases (Abalone, Quillback and Kelp) the ensemble models performed equally well or better than any of the individual models when evaluated using spatial block CV. And in those three cases, the ensemble models were still well within one standard deviation of the better performing models. While built using weighted means, the results of the ensembles is nevertheless in line with expectations that an equally weighted ensemble will be more accurate than at least half of the individual models (Araújo and New 2007).

We found similar patterns in the distributions of model performance statistics among GLM, BRT and ensemble models from all 8 species for which testing data was available (Figure 4.4). When compared across species, the HSI models had greater variation in performance metrics and scored lowest for all metrics except specificity (the correct prediction of absences), where the median was in line with the other modelling methods. Therefore, the HSI models performed similarly to GLM, BRT and ensemble models when predicting species absence, but performed relatively poorly when predicting species presence.

Spatially, at a broad scale, there is a high degree of agreement in the prediction of probability of occurrence between model types (Figure 4.5). However, differences are evident upon close examination where a patchiness in the predictions is seen to increase with model complexity from HSI to BRT. This pattern was evident across all species.

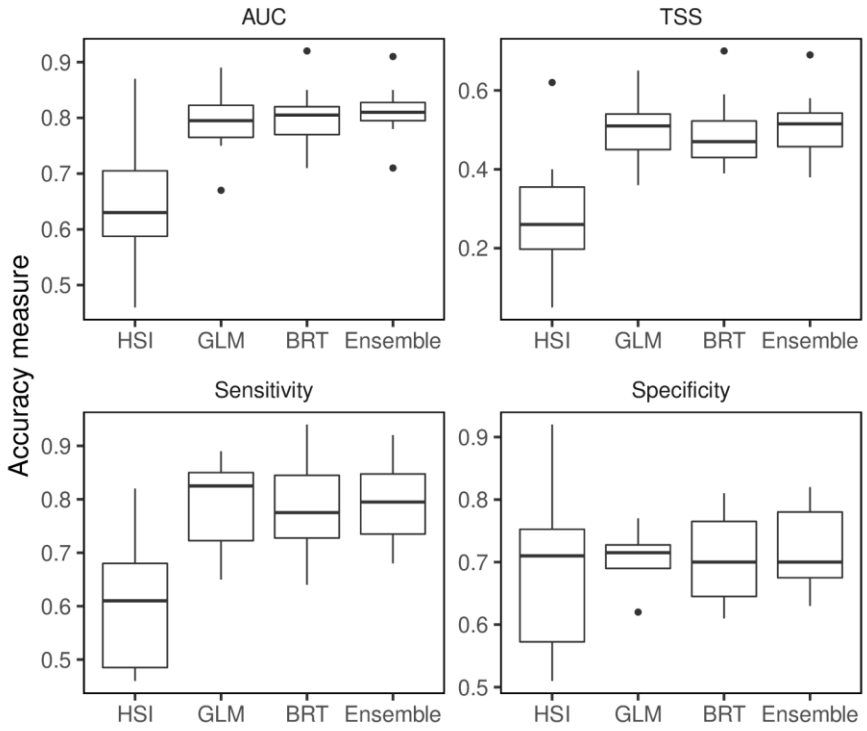

Figure 4.4. Boxplots showing the distribution of model performance metrics from the habitat suitability index (HSI), generalized linear (GLM), boosted regression tree (BRT), and ensemble models across all species models tested. AUC is the area under the receiver operator characteristic curve and TSS is the true skill statistic.

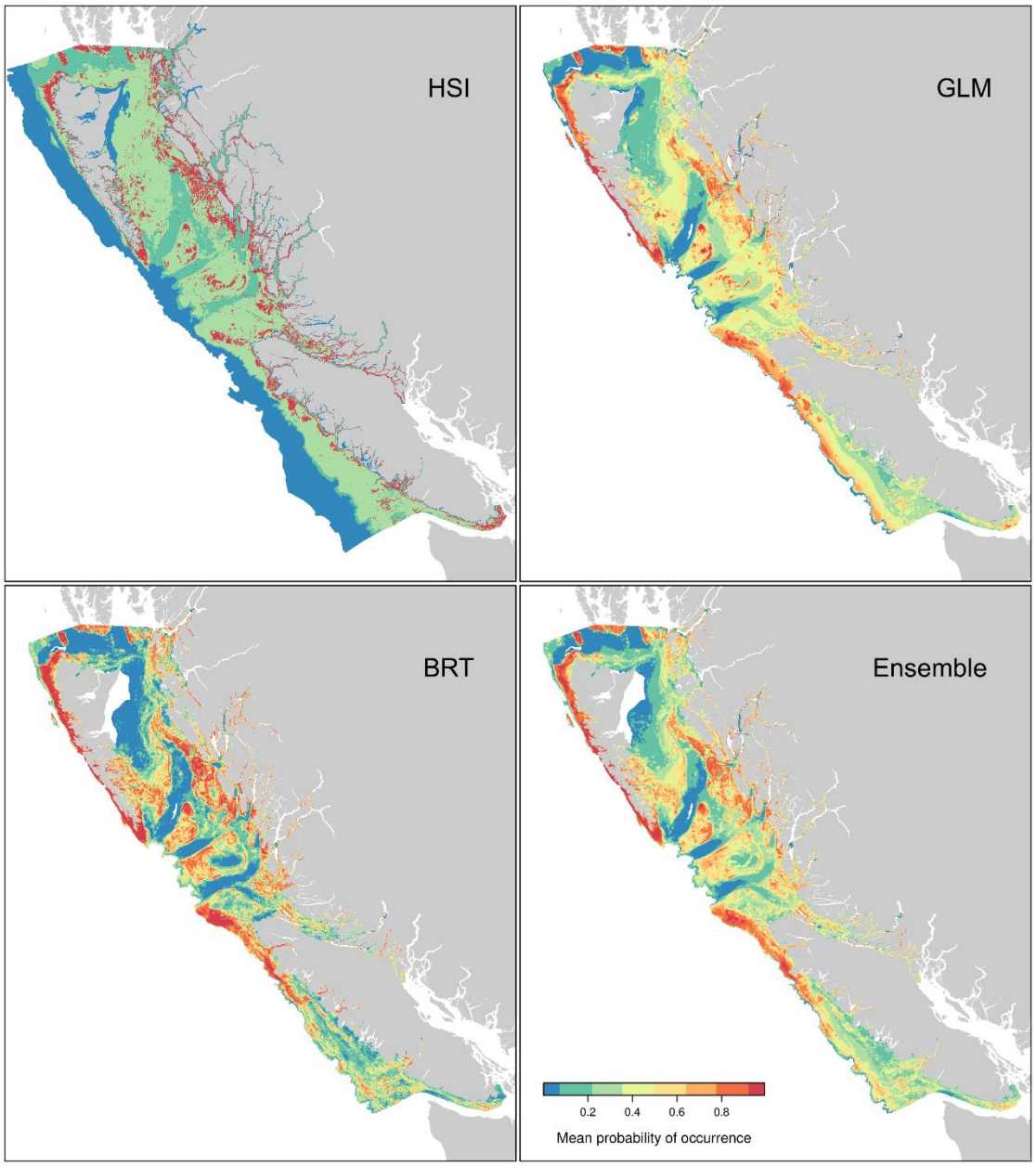

Figure 4.5. Species probability of occurrence predictions within the shelf study area from Quillback Rockfish habitat suitability index (HSI) model, generalized linear model (GLM), boosted regression tree (BRT) model, and ensemble model. With the exception of the HSI prediction, probability of occurrence is the mean prediction from five-fold cross-validation models. GLM, BRT and ensemble predictions are truncated to exclude areas of extrapolation in predictor space.

4.3.3 Model uncertainty

Following the methods outlined in the framework, four methods were used to examine model uncertainty for each species with ensemble models. For species for which only an HSI model was developed, we were not able to examine model uncertainty. In this application, we present model uncertainty as measured by standard deviation across model predictions (see Section 3.6, Figure 3.12), and, in Appendix A, as the difference between HSI model predictions and ensemble model predictions for each species.

Collectively, the BRT model predictions had slightly higher median uncertainty (standard deviation calculated from the eight species models with testing data) when compared to GLM model predictions (Figure 4.6). This is expected as BRT models are more flexible and fit closer to the training data than GLMs resulting in greater variation among prediction from CV models. The median uncertainty of the ensemble predictions is notably higher. This is the result of pooling the variance from both the CV models and the GLM and BRT models to incorporate both CV uncertainty and methodological uncertainty into a single estimate.

Overall ensemble model uncertainty, as measured by mean standard deviation across predictions, also varied among species models (Figure 4.7). Species models with higher AUC values, also tended to have lower mean uncertainty. Quillback Rockfish, the best performing ensemble model, had relatively low mean uncertainty while Dungeness Crab, the poorest performing ensemble model, had the highest mean uncertainty. Some species have relatively higher or lower mean uncertainty given their performance statistics (e.g., Abalone and Eelgrass). Eelgrass may have very low uncertainty because the predictions tend to be low on average due to relatively little potential Eelgrass habitat within the study area. Conversely, mean uncertainty may be higher than expected given model performance (e.g., Abalone) when non-stationarity exists within the study area. Although evidence of non-stationarity can be revealed by examining performance metrics calculated using test data created with spatial blocking, prediction uncertainty measured in this way can reveal additional variation between CV model predictions outside of sampled areas not capture in performance metrics.

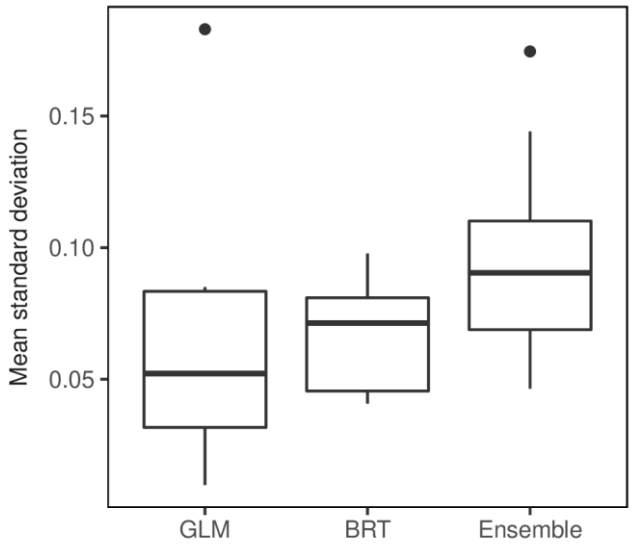

Figure 4.6. Distribution of uncertainty (as mean standard deviation across species predictions) for three modelling methods. Mean standard deviation of predictions is based on five-fold cross-validated model predictions from all species. Model methods included generalized linear model (GLM), boosted regression tree (BRT), and ensemble methods. Standard deviation of the ensemble models incorporates variance across GLM and BRT methods, as well as the individual CV models.

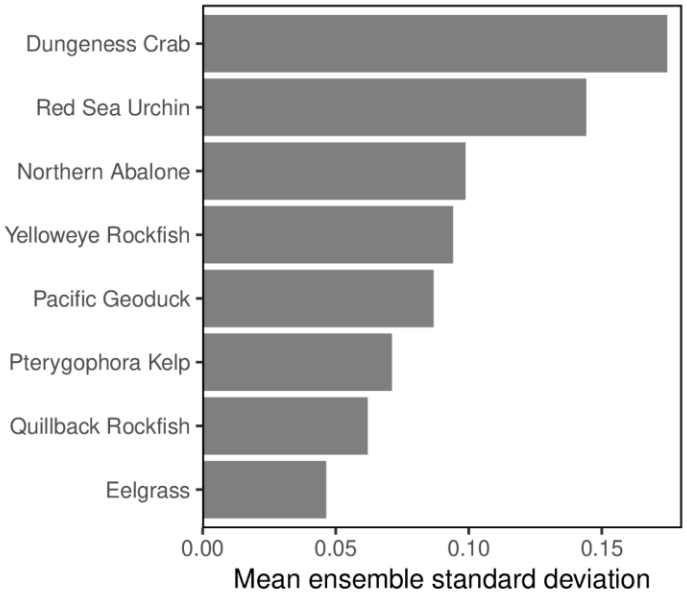

Figure 4.7. Uncertainty (as mean standard deviation from ensemble model predictions) for the eight species with sufficient occurrence data, ranked from highest to lowest.

4.3.4 Comparison of performance statistics

While AUC is a commonly used statistic, there is a diversity of literature examining different performance metrics (e.g., Lawson et al. 2014; Lobo et al. 2008). It is therefore desirable to compare the most common performance statistics used in the literature to determine whether any particular metric is a more reliable measure of performance.

A comparison of how the three threshold-dependent metrics (Kappa, TSS, and Accuracy) change across the range of threshold values from one example species (Figure 3.6) shows that Kappa and TSS track quite closely, staying within one standard deviation of each other. In contrast, Accuracy tended to have higher values and reached a plateau at higher threshold values. When threshold-dependent statistics from all species models were compared with the threshold-independent AUC statistic, a near perfect correlation (Spearman's rho = 0.98) was evident between TSS and AUC (Figure 4.8). In contrast, Accuracy and Kappa measures had a lower correlation with AUC (Spearman's rho = 0.77 and 0.60, respectively). On inspection, outliers (on the left side of the panel comparing AUC and Kappa, Figure 4.8) correspond to the models for two species with the lowest prevalence (Kelp and Eelgrass). This well documented effect of prevalence on Kappa (Allouche et al. 2006; Manel et al. 2001) is further illustrated by comparing Kappa to TSS for a representative low and high prevalence species (Figure 4.9). The effect is dramatic when prevalence is < 5%, and persists at prevalence levels of 25% indicating that a prevalence close to 50% is necessary to make Kappa comparable to TSS and AUC.

This comparison clearly shows how some performance statistics (i.e., AUC, TSS) performed more consistently then others (Accuracy, Kappa).

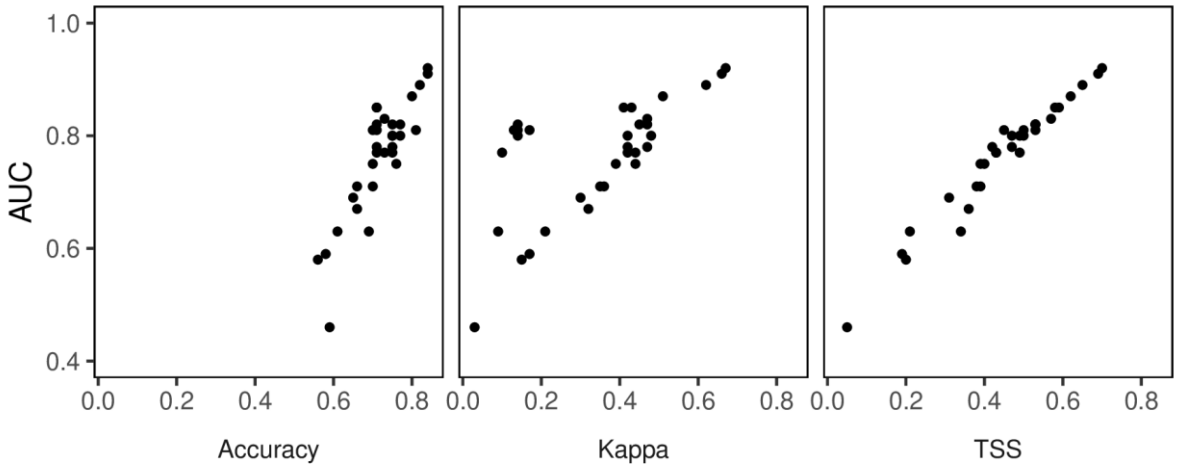

Figure 4.8. Correlation between the threshold-based predictors, Accuracy, Kappa, and TSS (true skill statistic), and the area under the receiver operator characteristic curve (AUC). All species and modelling methods were used for the comparison.

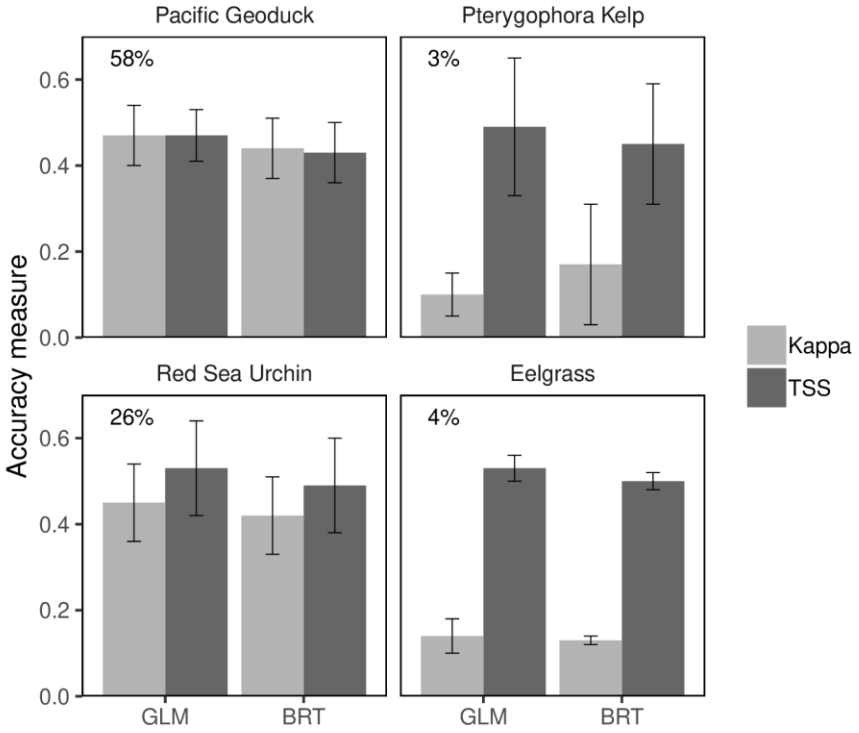

Figure 4.9. Mean performance for Kappa and TSS (true skill statistic) from five-fold cross-validation for example species with high and low prevalence. Validation metrics are shown separately for generalized linear (GLM) and boosted regression tree (BRT) models. Prevalence is reported in the upper left corner for each species. Error bars represent one standard deviation.

4.4 CONCLUSIONS

The twelve species distribution models presented here demonstrate how the framework presented in Section 3 can be effectively applied to create credible predictions. The resulting

probability of occurrence predictions will be used to inform emergency response in the event of a marine oil spill incident. For the eight species with adequate occurrence observations, model performance metrics and spatially explicit uncertainty measures provide managers with tools to assess confidence in the model predictions in a given area of interest. For the four species found to be data deficient, the preliminary envelope models provide a baseline from which data-driven models of distribution can be developed once observational data are available. By following a standardized model development process, the distribution models were produced efficiently while ensuring consistency across species. The use of a single, suite of predictor data (i.e., a consistent data model) was central to the ability of the framework to handle multiple species models in an efficient way. Additionally, during the process, the framework (along with the associated R scripts) facilitated updates when new occurrence data or improved predictors became available. The iterative approach to re-assessing model inputs for their suitability during the model interpretation phase and subsequently refining the models (see Figure 1.1) led to improved model performance in many cases and is a valuable feature of the framework.

This application also demonstrated the utility of a multiple model approach to SDM development. Building knowledge-based envelope models helped define the current ecological understanding of the twelve species and served as a benchmark against which the data-driven ensemble model predictions could be compared. Building ensemble model predictions allowed for uncertainty to be estimated by examining the variation across component models. The relative influence and marginal effects of predictors (see Appendix A) also helped inform reliability assessments for models of individual species.

This application was particularly successful for species of commercial or conservation interest, for which a large sample of suitable observations was available, although some species (e.g., Dungeness Crab), had observational data that was spatially biased toward more suitable habitat. For those species, additional observations obtained through a well-designed sampling program would improve the model predictions. For species with high-quality observational data (i.e., abundance data at appropriate resolution and extents with limited spatial sampling bias) and good association with the predictor variables developed here, models of species density are likely within reach and would provide additional information for oil spill response and other management needs.

5 RECOMMENDATIONS

The following recommendations were conceived during the development of the framework and the resulting application.

- Follow established modelling practice by clearly identifying model objectives. They inform the development of the data and statistical models and are central to interpreting model results, in part by allowing clear distinctions to be made between what is technically wrong (e.g., a poorly scaled predictor), and what is out of scope (e.g., juvenile life history stages), thereby reducing model misinterpretation.
- Build knowledge-based envelope models, regardless of the availability of observation data. Such models clarify the current ecological understanding the species, help identify predictors that are essential to effective habitat characterisation, and provide a means of calculating model uncertainty by comparing the knowledge-based to data-driven model predictions.
- Incorporate LEK wherever possible, to compliment scientific ecological knowledge. Such knowledge would be especially valuable in data poor situations, could be included in

envelope models, and combined as prior knowledge with data-driven approaches in a Bayesian framework.

- Examine the spatial and temporal coherence in resolution between the predictor and species occurrence data to assess the appropriateness of the data model and identify scale mismatches prior to model fitting.
- Consider how the quality of occurrence data influences model performance metrics. If confidence in the quality of the species data is low, high performance scores alone will not be indicative of a good model, as these metrics also reflect the quality of the occurrence data (i.e., if the input data is biased so will the evaluation). Different observational datasets can produce models with very similar performance metrics yet very different predictions of probability of occurrence across the study area.
- Consider reducing the sample size if a portion of the occurrence data (e.g., a particular survey or sampling method) lowers the overall data quality. There is often a trade-off between sample size and data quality, particularly if occurrence data are compiled from multiple sources. Fewer high quality data can produce a better model than more, low quality data.
- Employ spatial block CV to build models and evaluate performance. Effective blocking of observational data reduces the dependence between testing data and training data and produces model evaluation scores that are a more accurate measure of model transferability. To develop the broadest perspective on model performance, calculate metrics from both training and testing data to evaluate model fit and predictive performance.
- Examine validation plots such as those included in this framework (i.e., threshold plots, graphical contingency matrix and ROC curve) in addition to individual statistics to assess model performance. Graphical representations provide more information than single numbers, and are thus more informative, especially when comparing across multiple models.
- The threshold-independent AUC and threshold-dependent TSS statistics are recommended over the prevalence-sensitive Kappa and the less descriptive Accuracy statistics.
- If spatial autocorrelation remains in the residuals after model fitting, consider a spatial predictor such as an auto-covariate term to aid in model interpretation and understanding. Models with spatial predictors are not always appropriate for prediction, however comparing them to environmental-only models reveals the relative importance of environmental variables once spatial structure is accounted for and may help indicate missing processes by estimating unexplained spatial structure.
- Diagnose possible errors in the data and statistical models by looking for unexpected relationships in marginal effects and relative influence of predictors. Such plots can provide important insight into the structural uncertainty of the model and unaccounted for bias in the occurrence data.
- Create ensemble predictions from multiple models (e.g., from different model fitting methods) and calculate variation across source models to provide a spatially explicit measure of model uncertainty.
- Consider using independently collected data for additional model validation where practical.

6 GLOSSARY

This section contains a collection of terms whose definition is critical to understanding the contents of this document. We have included this glossary to provide readers a centralized reference to the technical terminology.

AUC: The Area Under the receiver-operator Curve. A common metric for estimating the performance of habitat suitability models.

AIC: The Akaike Information Criterion is a widely used measure of relative model performance intended to balance the trade-off between model fit to data, and parsimony. Given a particular correlative method and a set of observations and predictor data, it is used to select from a finite set of models, where the model with the lowest AIC is preferred. For statistical models, it is a function of degrees of freedom in the observational data.

Bias: The effect of non-random sampling of a variable. Biased data can obscure the pattern of interest and lead to incorrect inference.

BIC: Similar to the AIC, the Bayesian Information Criterion is also used to select among alternative models. Based in part on the likelihood function, it has somewhat larger penalty term than AIC, and is consequently more likely to select less complex models.

Coefficient of variation: The ratio of standard deviation to the mean.

Complexity: Managing complexity is a key part of model building. It has been variously described in the literature. Here, unless otherwise qualified, it refers to the complexity of model structure, and is equivalent to degrees of freedom used by the model.

Cross-validation (CV): Process of partitioning data into multiple (K) folds and repeatedly using K-1 folds as training data subsets for model fitting, and using the remaining fold for evaluation.

Deviance: Is a goodness-of-fit measure calculated from the residuals of a statistical model.

Ecosystem models: A broad class of models used to represent a particular aspect of an ecological system, usually in mathematical terms.

Habitat suitability index (HSI) model: A knowledge-based envelope model drawing on literature and expert consultation to describe species relationships with the environment.

Independent samples: A sample or collection of data is independent from another if there is no relationship between them that would allow inferences to be made about one from the other.

Local ecological knowledge (LEK): Place-based ecological knowledge held by a specific group of people related to living organisms and their relationship with the environment, including traditional ecological knowledge passed down through generations.

Model extents: The geographic space over which the model is constructed.

Presence-Absence model: These models rely on relatively certain absence data, obtained via monitoring or repeated visits. The reliability of absence data depends on the species' characteristics (e.g., biology, behaviour, history), their local abundance and ease of detection, and the survey design (Hirzel et al. 2006).

Presence-only model: A model that uses observations of presence only, making it difficult to assess a model's specificity. This challenge has been addressed in two ways: 1) use of pseudo-absences to allow the use of presence-absence frameworks; and 2) evaluation of predictions based on their deviance from random (Hirzel et al. 2006).

Prevalence: The proportion of presence or abundance values in dependent data set (requires absence data to be included in the data).

Sensitivity: The proportion of presence observations that are correctly predicted as presence

Spatial autocorrelation (SAC): Observations close in space are more likely to be similar than those further apart.

Species distribution model (SDM): A predictive model of the probability of species occurrence across space. While typically based on correlations between observations of species and habitat characteristics, species interactions are increasingly being considered.

Specificity: The proportion of absence observations that are correctly predicted as absence

Stationarity: A key assumption about the processes that underlie the relationships between species and their habitat. These processes are termed stationary when they do not vary across the study area, spatially and temporally. The likelihood of this assumption being false increases with model extents.

Threshold: The cut-off used to convert predictions from continuous (probability of occurrence) to binary (presence/absence).

Variance: A measure of the spread of a set of values around their mean.

7 ACKNOWLEDGMENTS

The authors would like to acknowledge the many people who contributed to the successful completion of this work. Anders Knudby, Chris Rooper and Javier Murillo-Perez provided thoughtful and knowledgeable reviews which led to important improvements to this manuscript. Matthew Grinnell made a significant contribution to the R scripts that build and evaluate the data-driven models. Leslie Barton, Michelle Bigg, James Boutillier, Dominique Bureau, Dan Curtis, Sarah Davies, Kieran Forge, Dana Haggarty, Christine Hansen, Elise Keppel, Joanne Lessard, Dan Leus, Tammy Norgard and Emily Rubidge shared their ecological knowledge on which the HSI models were based. Sean Anderson, Katherine Bannar-Martin, Mauricio Carrasquilla, Lisa Lacko, Shannon Obradovich, Maria Surry, Katie Gale and Kim Conway provided advice and support related to species and environmental data. Lastly, the authors gratefully acknowledge everyone who has contributed to the collection and management of the numerous sources of species and environmental data used here without which none of this work would have been possible.

8 REFERENCES

Allouche, O., Tsoar, A., and Kadmon, R. 2006. Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). J. Appl. Ecol. **43**: 1223-1232. doi:10.1111/j.1365-2664.2006.01214.x.

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. doi:10.1016/j.dsr.2016.07.006.

Araújo, M.B., and New, M. 2007. Ensemble forecasting of species distributions. Trends Ecol. Evol. **22**(1): 42-47.

Araújo, M.B., and Peterson, A.T. 2012. Uses and misuses of bioclimatic envelope modeling. Ecology **93**(7): 1527-1539. Araújo, M.B., Anderson, R.P., Barbosa, A.M., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., and Naimi, B. 2019. Standards for distribution models in biodiversity assessments. Science advances **5**(1): eaat4858.

Augustin, A.N.H., Mugglestone, M.A., and Buckland, S.T. 1996. An Autologistic Model for the Spatial Distribution of Wildlife. J. Appl. Ecol. **33**: 339-347.

Austin, M. 2007. Species distribution models and ecological theory: a critical assessment and some possible new approaches. Ecol. Model. **200**(1): 1-19.

Austin, M.P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecol. Model. **157**: 101-118.

Bahn, V., and McGill, B.J. 2007. Can niche-based distribution models outperform spatial interpolation? Global Ecol. Biogeogr. **16**: 733-742. doi:10.1111/j.1466-8238.2007.00331.x.

Bahn, V., and McGill, B.J. 2013. Testing the predictive performance of distribution models. Oikos **122**(3): 321-331. doi:10.1111/j.1600-0706.2012.00299.x.

Bardos, D.C., Guillera-Arroita, G., and Wintle, B.A. 2015. Valid auto-models for spatially autocorrelated occupancy and abundance data. Methods in Ecology and Evolution **6**: 1137-1149. doi:10.1111/2041-210X.12402.

Barry, S., and Elith, J. 2006. Error and uncertainty in habitat models. J. Appl. Ecol. **43**(3): 413-423.

Barton, K. 2018. MuMin: Multi-model inference. R package version 1.15. 6.; 2016.

Beger, M., and Possingham, H.P. 2008. Environmental factors that influence the distribution of coral reef fishes: modeling occurrence data for broad-scale conservation and management. Mar. Ecol. Prog. Ser. **361**: 1-13.

Belisle, A.C., Asselin, H., LeBlanc, P., and Gauthier, S. 2018. Local knowledge in ecological modeling. Ecol. Soc. **23**(2): 14.

Bell, D.M., and Schlaepfer, D.R. 2016. On the dangers of model complexity without ecological justification in species distribution modeling. Ecol. Model. **330**: 50-59.

Bini, L.M., Diniz-Filho, J.A.F., Rangel, T.F., Akre, T.S., Albaladejo, R.G., Albuquerque, F.S., Aparicio, A., Araujo, M.B., Baselga, A., and Beck, J. 2009. Coefficient shifts in geographical ecology: an empirical evaluation of spatial and non-spatial regression. Ecography **32**(2): 193-204.

Breen, P., and Adkins, B. 1979. A survey of abalone populations on the east coast of the Queen Charlotte Islands, August 1978. Fish. Mar. Serv. Manuscr. Rep **1490**: 125.

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

Brooks, R.P. 1997. Improving Habitat Suitability Index Models. Wildlife Society Bulletin (1973-2006) **25**(1): 163-167.

Burnham, K.P., and Anderson, D.R. 2004. Model selection and multi-model inference: a Practical Information-Theoretic Approach. Second ed. Springer-Verlag, New York.

Burrows, M.T. 2012. Influences of wave fetch, tidal flow and ocean colour on subtidal rocky communities. Mar. Ecol. Prog. Ser. **445**: 193-207.

Carignan, K., Eakins, B., Love, M., Sutherland, M., and McLean, S. 2013. Bathymetric Digital Elevation Model of British Columbia, Canada: Procedures, Data Sources, and Analysis. NOAA National Geophysical Data Center (NGDC). 8 pp.

Carrasquilla-Henao, M., Yamanaka, K.L., Haggarty, D., and Juanes, F. 2018. Predicting important rockfish (Sebastes spp.) habitat from large-scale longline surveys for southern British Columbia, Canada Can. J. Fish. Aquat. Sci. **13**: 1-13. doi:10.1139/cjfas-2017-0458.

Cheung, W.W.L., Lam, V.W.Y., Sarmiento, J.L., Kearney, K., Watson, R., and Pauly, D. 2009. Projecting global marine biodiversity impacts under climate change scenarios. Fish Fish. **10**: 235-251.

Colding, J., and Folke, C. 2001. Social taboos: “invisible” systems of local resource management and biological conservation. **11**(2): 584-600.

Dallas, T.A., and Hastings, A. 2018. Habitat suitability estimated by niche models is largely unrelated to species abundance. Global Ecol. Biogeogr. **27**(12): 1448-1456.

Davies, S.C., Gregr, E.J., Lessard, J., Bartier, P., and Wills, P. 2019. Bathymetric elevation models for ecological analyses in Pacific Canadian coastal waters. Can. Tech. Rep. Fish. Aquat. Sci. 3321: vi + 38 p.

De'Ath, G., and Fabricius, K.E. 2000. Classification and Regression Trees: A Powerful Yet Simple Technique for Ecological Data Analysis. Ecology **81**: 3178-3192.

DFO. 2009. Development of a framework and principles for the biogeographic classification of Canadian marine areas. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2009/056.

Dormann, C.F., Purschke, O., Márquez, J.R.G., Lautenbach, S., and Schröder, B. 2008. Components of uncertainty in species distribution analysis: a case study of the great grey shrike. Ecology **89**(12): 3371-3386.

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.

Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, D.W., Peres-Neto, P.R., Ohlemüller, R., Reineking, B., Kühn, I., Schröder, B., Schurr, F.M., and Wilson, R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. Ecography **30**: 609-628. doi:10.1111/j.2007.0906-7590.05171.x.

Du Preez, C. 2015. A new arc–chord ratio (ACR) rugosity index for quantifying three-dimensional landscape structural complexity. Landscape Ecol. **30**(1): 181-192.

Du Preez, C., Curtis, J.M., and Clarke, M.E. 2016. The structure and distribution of benthic communities on a shallow seamount (Cobb Seamount, Northeast Pacific Ocean). PloS one **11**(10): e0165513.

Eger, A.M., Curtis, J.M., Fortin, M.-J., Côté, I.M., and Guichard, F. 2016. Transferability and scalability of species distribution models: a test with sedentary marine invertebrates. Can. J. Fish. Aquat. Sci. **74**(5): 766-778.

Elith, J., and Graham, C.H. 2009. Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. Ecography **32**(1): 66-77.

Elith, J., and Leathwick, J.R. 2009. Species distribution models: ecological explanation and prediction across space and time. Annu. Rev. Ecol., Evol. Syst. **40**: 677-697.

Elith, J., Burgman, M.A., and Regan, H. 2002. Mapping epistemic uncertainties and vague concepts in predictions of species distribution. Ecol. Model. **157**: 313-329.

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

Elith, J., Ferrier, S., Huettmann, F., and Leathwick, J. 2005. The evaluation strip: A new and robust method for plotting predicted responses from species distribution models. Ecol. Model. **186**(280-289).

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

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.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberon, J., Williams, S., Wisz, M.S., and Zimmermann, N.E. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**: 129-151.

Ericksen, P., and Woodley, E. 2005. Using multiple knowledge systems: benefits and challenges. *In* Ecosystems and human well-being: multiscale assessments. Island Press, Washington, DC, USA: 85-117.

ESRI. 2019. ArcGIS. Environmental Systems Research Institute, Redlands, CA.

Evans, M.R., Grimm, V., Johst, K., Knuuttila, T., De Langhe, R., Lessells, C.M., Merz, M., O'Malley, M.A., Orzack, S.H., and Weisberg, M. 2013. Do simple models lead to generality in ecology? Trends Ecol. Evol. **28**(10): 578-583.

Foster, S.D., Dunstan, P.K., Althaus, F., and Williams, A. 2015. The cumulative effect of trawl fishing on a multispecies fish assemblage in south-eastern Australia. J. Appl. Ecol. **52**(1): 129-139.

Fourcade, Y., Besnard, A.G., and Secondi, J. 2018. Paintings predict the distribution of species, or the challenge of selecting environmental predictors and evaluation statistics. Global Ecol. Biogeogr. **27**(2): 245-256.

Freeman, E.A., and Moisen, G.G. 2008. A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. Ecol. Model. **217**(1-2): 48-58.

Friedman, J.H., and Meulman, J.J. 2003. Multiple additive regression trees with application in epidemiology. Stat. Med. **22**: 1365-1381. doi:10.1002/sim.1501.

Gale, K.S.P., Frid, A., Lee, L., McCarthy, J., Robb, C., Rubidge, E., Steele, J., and Curtis, J.M.R. 2019. A framework for identification of ecological conservation priorities for Marine Protected Area network design and its application in the Northern Shelf Bioregion. DFO Can. Sci. Advis. Sec. Res. Doc. 2018/055. viii + 186 p..

Galipaud, M., Gillingham, M.A.F., David, M., and Dechaume-Moncharmont, F.X. 2014. Ecologists overestimate the importance of predictor variables in model averaging: A plea for cautious interpretations. Methods in Ecology and Evolution **5**: 983-991. doi:10.1111/2041-210X.12251.

García-Callejas, D., and Araújo, M.B. 2016. The effects of model and data complexity on predictions from species distributions models. Ecol. Model. **326**: 4-12.

Garrison, B., and Lupo, T. 2002. Accuracy of bird range maps based on habitat maps and habitat relationships models. *In* Predicting species occurrences. Issues of scale and accuracy. *Edited by* J. Scott and P. Heglund and M. Morisson and J. Haufler and M. Raphael and W. Wall and F. Samson. Island Press, Washington. pp. 367-375.

Greenwell, B., Boehmke, B., and Cunningham, J. 2018. gbm: Generalized Boosted Regression Models. R package version 2.1.4. .

Gregr, E.J. 2011. Insights into North Pacific right whale *Eubalaena japonica* habitat from historic whaling records. Endangered Species Research **15**: 223-239.

Gregr, E.J. 2012. BC_EEZ_100m: A 100 m raster of the Canadian Pacific Exclusive Economic Zone. SciTech Environmental Consulting, Vancouver BC.

Gregr, E.J. 2014. Fetch Geometry Calculator Version 1.0 – User Guide. SciTech Environmental Consulting, Vancouver, BC. 5 pp.

Gregr, E.J., and Chan, K.M.A. 2014. Leaps of Faith: How implicit assumptions compromise the utility of ecosystem models for decision-making. Bioscience **65**(1): 43-54.

Gregr, E.J., and Haggarty, D. 2017. Background substrate and the integration of nearshore and deep water classifications (Draft). SciTech Environmental Consulting, Vancouver, BC. 26 pp.

Gregr, E.J., Lessard, J., and Harper, J. 2013. A spatial framework for representing nearshore ecosystems. Prog. Oceanogr. **115**: 189-201.

Gregr, E.J., Palacios, D.M., Thompson, A., and Chan, K.M. 2018. Why less complexity produces better forecasts: An independent data evaluation of kelp habitat models. Ecography **42**(3):428-43. doi:10.1111/ecog.03470.

Grimm, V., Augusiak, J., Focks, A., Franke, B.M., Gabsi, F., Johnston, A.S.A., Liu, C., Martin, B.T., Meli, M., Radchuk, V., Thorbek, P., and Railsback, S.F. 2014. Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. Ecol. Model. **280**: 129-139.

Grömping, U. 2015. Relative Importance for Linear Regression in R: The Package relaimpo. Journal of Statistical Software **17**. doi:10.18637/jss.v017.i01.

Guinotte, J.M., and Davies, A.J. 2014. Predicted deep-sea coral habitat suitability for the U.S. West Coast. PLoS ONE **9**. doi:10.1371/journal.pone.0093918.

Guisan, A., and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. Ecol. Model. **135**: 147-186.

Guisan A, Thomas C. Edwards J, Hastie T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157:89-100.

Guisan, A., Thuiller, W., and Zimmermann, N.E. 2017. Habitat suitability and distribution models: with applications in R. Cambridge University Press.

Hannah, L., St. Germain, C., Jeffery, S., Patton, S., and O, M. 2017. Application of a framework to assess vulnerability of biological components to ship-source oil spills in the marine environment in the Pacific Region. DFO Can. Sci. Advis. Sec. Res. Doc. 2017/057. ix + 145 p.

Hastie, T., Tibshirani, R., and Friedman, J. 2009. The Elements of Statistical Learning. Elements **1**: 337-387. doi:10.1007/b94608.

Hawkins, B.A. 2012. Eight (and a half) deadly sins of spatial analysis. J. Biogeogr. **39**(1): 1-9.

Hawkins, D.M., Basak, S.C., and Mills, D. 2003. Assessing model fit by cross-validation. Journal of Chemical Information and Computer Sciences **43**: 579-586. doi:10.1021/ci025626i.

Hepinstall, J., Krohn, W., and Sader, S. 2002. Effects of niche width on the performance and agreement of avian habitat models. *In* Predicting species occurrences. Issues of scale and accuracy. *Edited by* J. Scott and P. Heglund and M. Morisson and J. Haufler and M. Raphael and W. Wall and F. Samson. Island Press, Washington. pp. 593-606.

Hernandez, P.A., Graham, C.H., Master, L.L., and Albert, D.L. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography **29**: 773-785.

Hijmans, R.J. 2012. Cross-validation of species distribution models: Removing spatial sorting bias and calibration with a null model. Ecology **93**: 679-688. doi:10.1890/11-0826.1.

Hijmans, R.J., Phillips, S., Leathwick, J., and Elith, J. 2017. dismo: Species Distribution Modeling. R package version 1.1-4.

Hirzel, A.H., and Le Lay, G. 2008. Habitat suitability modelling and niche theory. J. Appl. Ecol. **45**(5): 1372-1381.

Hirzel, A.H., Hausser, J., Chessel, D., and Perrin, N. 2002. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? Ecology **83**(7): 2027-2036.

Hirzel, A.H., Lay, G.L., Helfer, V., Randin, C., and Guisan, A. 2006. Evaluating the ability of habitat suitability models to predict species presences. Ecol. Model. **199**: 142-152.

Hothorn, T., Müller, J., Schröder, B., Kneib, T., and Brandl, R. 2011. Decomposing environmental, spatial, and spatiotemporal components of species distributions. Ecol. Monogr. **81**: 329-347. doi:10.1890/10-0602.1.

Hu, C., Lee, Z., and Franz, B. 2012. Chlorophyll a algorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. J. Geophys. Res. (C Oceans) **117**: 1-25. doi:10.1029/2011JC007395.

Hughes, J., Haran, M., and Caragea, P.C. 2011. Autologistic models for binary data on a lattice. Environmetrics **22**: 857-871. doi:10.1002/env.1102.

Iturbide, M., Bedia, J., and Gutiérrez, J. 2018. Tackling Uncertainties of Species Distribution Model Projections with Package mopa. R Journal **10**(1).

Jones-Farrand, D.T., Fearer, T.M., Thogmartin, W.E., Thompson, F.R.I., Nelson, M.D., and Tirpak, J.M. 2011. Comparison of statistical and theoretical habitat models for conservation planning: the benefit of ensemble prediction. Ecol. Appl. **21**: 2269-2282. doi:10.1890/10-1047.1.

Jones, C.C., Acker, S.A., and Halpern, C.B. 2010. Combining local-and large-scale models to predict the distributions of invasive plant species. Ecol. Appl. **20**(2): 311-326.

Jones, M.C., and Cheung, W.W.L. 2015. Multi-model ensemble projections of climate change effects on. ICES J. Mar. Sci. **72**: 741-752.

Kaschner, K., Watson, R., Trites, A.W., and Pauly, D. 2006. Mapping world-wide distributions of marine mammal species using a relative environmental suitability (RES) model. Mar. Ecol. Prog. Ser. **316**: 285-310.

Keitt, T.H., Bjørnstad, O.N., Dixon, P.M., and Citron-Pousty, S. 2002. Accounting for spatial pattern when modeling organism-environment interactions. Ecography **25**(5): 616-625.

Kühn, I., and Dormann, C.F. 2012. Less than eight (and a half) misconceptions of spatial analysis. J. Biogeogr. **39**(5): 995-998.

Landis, J.R., and Koch, G.G. 1977. The measurement of observer agreement for categorical data. Biometrics: 159-174.

Lawson, C.R., Hodgson, J.A., Wilson, R.J., and Richards, S.A. 2014. Prevalence, thresholds and the performance of presence–absence models. Methods in Ecology and Evolution **5**(1): 54-64.

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. Mar. Ecol. Prog. Ser. **321**: 267-281.

Leathwick, J., Elith, J., Chadderton, W., Rowe, D., and Hastie, T. 2008. Dispersal, disturbance and the contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species. J. Biogeogr. **35**(8): 1481-1497.

Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm? Ecology **74**(6): 1659-1673.

Legendre, P., and Fortin, M. 1989. Spatial pattern and ecological analysis. Plant Ecol. **80**: 107-138.

Lessard, J., and Campbell, A. 2007. Describing northern abalone, *Haliotis kamtschatkana*, habitat: focusing rebuilding efforts in British Columbia, Canada. J. Shellfish Res. **26**(3): 677-686.

Levin, S.A. 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. Ecology **73**(6): 1943-1967.

Levins, R. 1966. The strategy of model building in population biology. Am. Sci. **54**(4): 421-431.

Li, J., Heap, A.D., Potter, A., and Daniell, J.J. 2011. Application of machine learning methods to spatial interpolation of environmental variables. Environ. Model. Software **26**(12): 1647-1659.

Liaw, A., and Wiener, M. 2002. Classification and regression by randomForest. R news **2**(3): 18-22.

Lichstein, J.W., Simons, T.R., Shriner, S.A., Franzreb, K.E., Monographs, S.E., and Aug, N. 2002. Spatial autocorrelation and autoregressive models in ecology. Ecol. Monogr. **72**: 445-463. doi:10.1890/0012-9615(2002)072[0445:saaami]2.0.co;2.

Link, J.S., Ihde, T.F., Harvey, C.J., Gaichas, S.K., Field, J.C., Brodziak, J.K.T., Townsend, H.M., and Peterman, R.M. 2012. Dealing with uncertainty in ecosystem models: The paradox of use for living marine resource management. Prog. Oceanogr. **102**: 102-114.

Link, W.A., and Barker, R.J. 2006. Ecological Society of America Model Weights and the Foundations of Multimodel Inference. Ecology **87**: 2626-2635.

Lobo, J.M., Jiménez-Valverde, A., and Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology & Biogeography **17**: 145-151.

Lobo, J.M., Jiménez-Valverde, A., and Hortal, J. 2010. The uncertain nature of absences and their importance in species distribution modelling. Ecography **33**(1): 103-114.

McPherson, J.M., and Jetz, W. 2007. Effects of species' ecology on the accuracy of distribution models. Ecography **30**(1): 135-151.

MacArthur, R. 1955. Fluctuations of animal populations and a measure of community stability. Ecology **36**(3): 533-536.

Manel, S., Williams, H.C., and Ormerod, S.J. 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. J. Appl. Ecol. **38**: 921-931.

Martin, M.P., Orton, T.G., Lacarce, E., Meersmans, J., Saby, N.P.A., Paroissien, J.B., Jolivet, C., Boulonne, L., and Arrouays, D. 2014. Evaluation of modelling approaches for predicting the spatial distribution of soil organic carbon stocks at the national scale. Geoderma **223-225**: 97-107. doi:10.1016/j.geoderma.2014.01.005.

Masson, D., and Fine, I. 2012. Modeling seasonal to interannual ocean variability of coastal British Columbia. J. Geophys. Res. (C Oceans) **117**(C10): C10019:10011-10014. doi:10.1029/2012jc008151.

Maxwell, S.M., Hazen, E.L., Lewison, R.L., Dunn, D.C., Bailey, H., Bograd, S.J., Briscoe, D.K., Fossette, S., Hobday, A.J., and Bennett, M. 2015. Dynamic ocean management: Defining and conceptualizing real-time management of the ocean. Mar. Policy **58**: 42-50.

Merckx, B., Steyaert, M., Vanreusel, A., Vincx, M., and Vanaverbeke, J. 2011. Null models reveal preferential sampling, spatial autocorrelation and overfitting in habitat suitability modelling. Ecol. Model. **222**(3): 588-597.

Merow, C., Smith, M.J., Edwards, T.C., Guisan, A., McMahon, S.M., Normand, S., Thuiller, W., Wüest, R.O., Zimmermann, N.E., and Elith, J. 2014. What do we gain from simplicity versus complexity in species distribution models? Ecography **37**(12): 1267-1281.

Oppel, S., Gardner, B., O'Connell, A.F., Louzao, M., Miller, P.I., Meirinho, A., and Ramírez, I. 2011. Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. Biol. Conserv. **156**: 94-104. doi:10.1016/j.biocon.2011.11.013.

Ord, J.K. 1979. Time-series and spatial patterns in ecology. *In* Spatial and Temporal Analysis in Ecology. *Edited by* R.M. Cormack and J.K. Ord. International Cooperative Publishing House, Fairland, Maryland, USA.

Pearce, J., and Ferrier, S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol. Model. **133**: 225–245.

Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Segurado, P., Dawson, T.P., and Lees, D.C. 2006. Model-based uncertainty in species range prediction. J. Biogeogr. **33**(10): 1704-1711.

Phillips, S.J., and Elith, J. 2010. POC plots: calibrating species distribution models with presence-only data. Ecology **91**(8): 2476-2484.

Planque, B., Loots, C., Petitgas, P., LindstrøM, U., and Vaz, S. 2011. Understanding what controls the spatial distribution of fish populations using a multi-model approach. Fish. Oceanogr. **20**(1): 1-17.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Raleigh, R.F., Miller, W.J., and Nelson, P.C. 1986. Habitat suitability index models and instream flow suitability curves: Chinook Salmon. U.S. Fish And Wildlife Service. Biological Report 82(10.122).

Randin, C.F., Dirnböck, T., Dullinger, S., Zimmermann, N.E., Zappa, M., and Guisan, A. 2006. Are niche-based species distribution models transferable in space? J. Biogeogr. **33**(10): 1689-1703.

Ridgeway, G. 2007. Generalized Boosted Models: A Guide to the gbm Package. *In* R package vignette.

Roberts, D.R., Bahn, V., Ciuti, S., Wintle, B.A., Guillera-Arroita, G., Elith, J., Warton, D.I., Hartig, F., Dormann, C.F., Lahoz-Monfort, J.J., Hauenstein, S., Thuiller, W., Schröder, B., and Boyce, M.S. 2016. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography **40**: 913-929. doi:10.1111/ecog.02881.

Robinson, N.M., Lundquist, C.J., Sutherland, J.E., Nelson, W.A., and Costello, M.J. 2017. A Systematic Review of Marine-Based Species Distribution Models (SDMs) with Recommendations for Best Practice. Frontiers in Marine Science **4**: 1-11. doi:10.3389/fmars.2017.00421.

Rooper, C.N., Zimmermann, M., and Prescott, M.M. 2017. Comparison of modeling methods to predict the spatial distribution of deep-sea coral and sponge in the Gulf of Alaska. Deep Sea Res. (I Oceanogr. Res. Pap.) **126**: 148-161.

Santoul, F., Mengin, N., Cereghino, R., Figuerola, J., and Mastrorillo, S. 2005. Environmental factors influencing the regional distribution and local density of a small benthic fish: the stoneloach (*Barbatula barbatula*). Hydrobiologia **544**(1): 347-355.

Sappington, J.M., Longshore, K.M., and Thompson, D.B. 2007. Quantifying landscape ruggedness for animal habitat analysis: a case study using bighorn sheep in the Mojave Desert. The Journal of wildlife management **71**(5): 1419-1426.

Schwarz, G. 1978. Estimating the dimension of a model. The annals of statistics **6**(2): 461-464.

Segurado, P., and Araújo, M.B. 2004. An evaluation of methods for modelling species distributions. J. Biogeogr. **31**: 1555-1568.

Silvert, W.L. 1981. Principles of ecosystem modelling. *In* Analysis of Marine Ecosystems. *Edited by* A.R. Longhurst. Academic Press, New York. pp. 651-676.

Soberón, J., and Peterson, A.T. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodiversity Informatics **2**: 1-10.

Soberón, J., and Nakamura, M. 2009. Niches and distributional areas: concepts, methods, and assumptions. PNAS **106**(Supplement 2): 19644-19650.

Stockwell, D.R., and Peterson, A.T. 2002. Effects of sample size on accuracy of species distribution models. Ecol. Model. **148**(1): 1-13.

Tabachnick, B.G., Fidell, L.S., and Ullman, J.B. 2007. Using multivariate statistics. Pearson, Boston, MA.

Tien, N., Craeymeersch, J., Van Damme, C., Couperus, A., Adema, J., and Tulp, I. 2017. Burrow distribution of three sandeel species relates to beam trawl fishing, sediment composition and water velocity, in Dutch coastal waters. J. Sea Res. **127**: 194-202.

Trachsel, M., and Telford, R.J. 2016. Technical note: Estimating unbiased transfer-function performances in spatially structured environments. Climate of the Past **12**: 1215-1223. doi:10.5194/cp-12-1215-2016.

Tuanmu, M.N., Viña, A., Roloff, G.J., Liu, W., Ouyang, Z., Zhang, H., and Liu, J. 2011. Temporal transferability of wildlife habitat models: implications for habitat monitoring. J. Biogeogr. **38**(8): 1510-1523.

USFWS. 1981. Standards for the development of habitat suitability index models. U.S. Department of Interior, Fish and Wildlife Service, Division of Ecological Services ESM 103. 162 + viii pp.

Valavi, R., Elith, J., Lahoz-Monfort, J.J., and Guillera-Arroita, G. 2018. blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods in Ecology and Evolution: 1-8. doi:10.1111/2041-210X.13107.

Walbridge, S., Slocum, N., Pobuda, M., and Wright, D. 2018. Unified geomorphological analysis workflows with benthic terrain modeler. Geosciences **8**(3): 94.

Wells, R.D., Boswell, K.M., Cowan Jr., J.H., and Patterson III, W.F. 2008. Size selectivity of sampling gears targeting red snapper in the northern Gulf of Mexico. Fisheries Research **89**(3): 294-299.

Wenger, S.J., and Olden, J.D. 2012. Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods in Ecology and Evolution **3**(2): 260-267.

Westerbom, M., and Jattu, S. 2006. Effects of wave exposure on the sublittoral distribution of blue mussels *Mytilus edulis* in a heterogeneous archipelago. Mar. Ecol. Prog. Ser. **306**: 191-200.

Wiens, J.A. 1989. Spatial scaling in ecology. Funct. Ecol. **3**: 385-397.

Wilson, E.O., and MacArthur, R.H. 1967. The theory of island biogeography. Princeton, NJ.

Yates, K.L., Bouchet, P.J., Caley, M.J., Mengersen, K., Randin, C.F., Parnell, S., Fielding, A.H., Bamford, A.J., Ban, S., and Barbosa, A.M. 2018. Outstanding challenges in the transferability of ecological models. Trends Ecol. Evol. **33**(10):790-802.

APPENDIX A. PREDICTIONS OF TWELVE SPECIES DISTRIBUTIONS FROM THE APPLICATION OF THE FRAMEWORK

This appendix presents the results from applying the framework presented in Section 4 to twelve species. For each species, we present figures showing:

1. the distribution of the presence and absence observations used for modelling,
2. predictions of species distribution and their related uncertainty, and
3. the relative influence and marginal effects of the environmental predictor variables.

Uncertainty is presented spatially by showing the difference between the knowledge-based habitat suitability index (HSI) model prediction and the ensemble model prediction and by showing the variation across multiple ensemble predictions using standard deviation (see Section 3.6). Each ensemble model combined predictions from a generalized linear model (GLM) and a boosted regression tree (BRT) model.

For species without appropriate data for building correlative models, ochre sea star (*Pisaster ochraceus*), blue mussel complex (*Mytilus edulis*, *M. trossulus*, and *M. galloprovincialis*), littleneck clam (*Leukoma staminea*), and orange sea pen (*Ptilosarcus gurneyi*), a single HSI model is presented with no uncertainty measure.

We cannot display presence observations or predicted distributions for Northern Abalone due to its Species at Risk (SARA) status. Restrictions around sharing information on known Abalone habitat are in place to help protect recovering Abalone populations. See Figure 4.3 for the relative influence and marginal effects relationships.

Pacific Geoduck

Panopea generosa

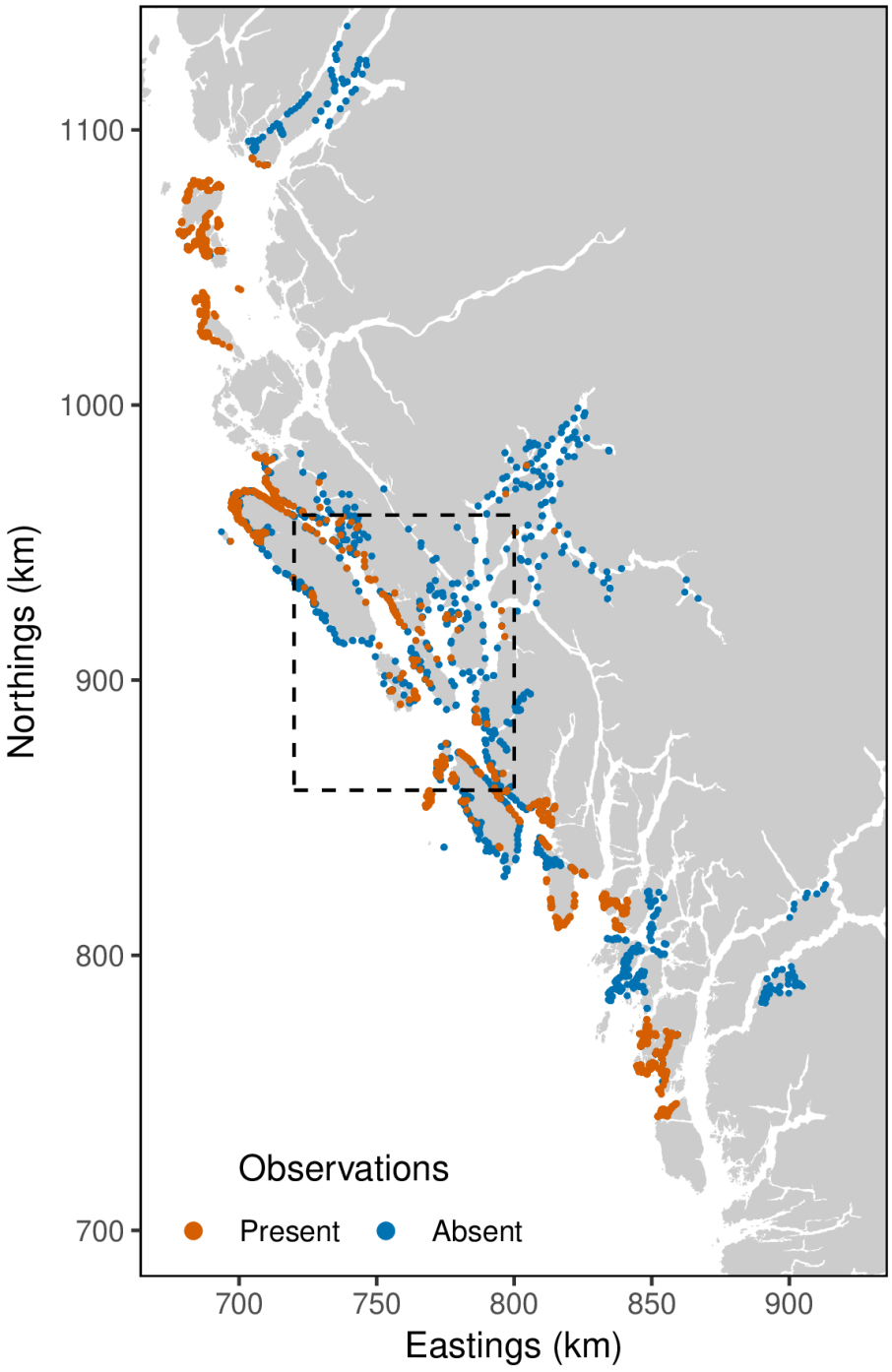

Figure A.1. Geoduck presence and absence observations within the nearshore study area. Dashed box shows the area displayed in Figure A.2.

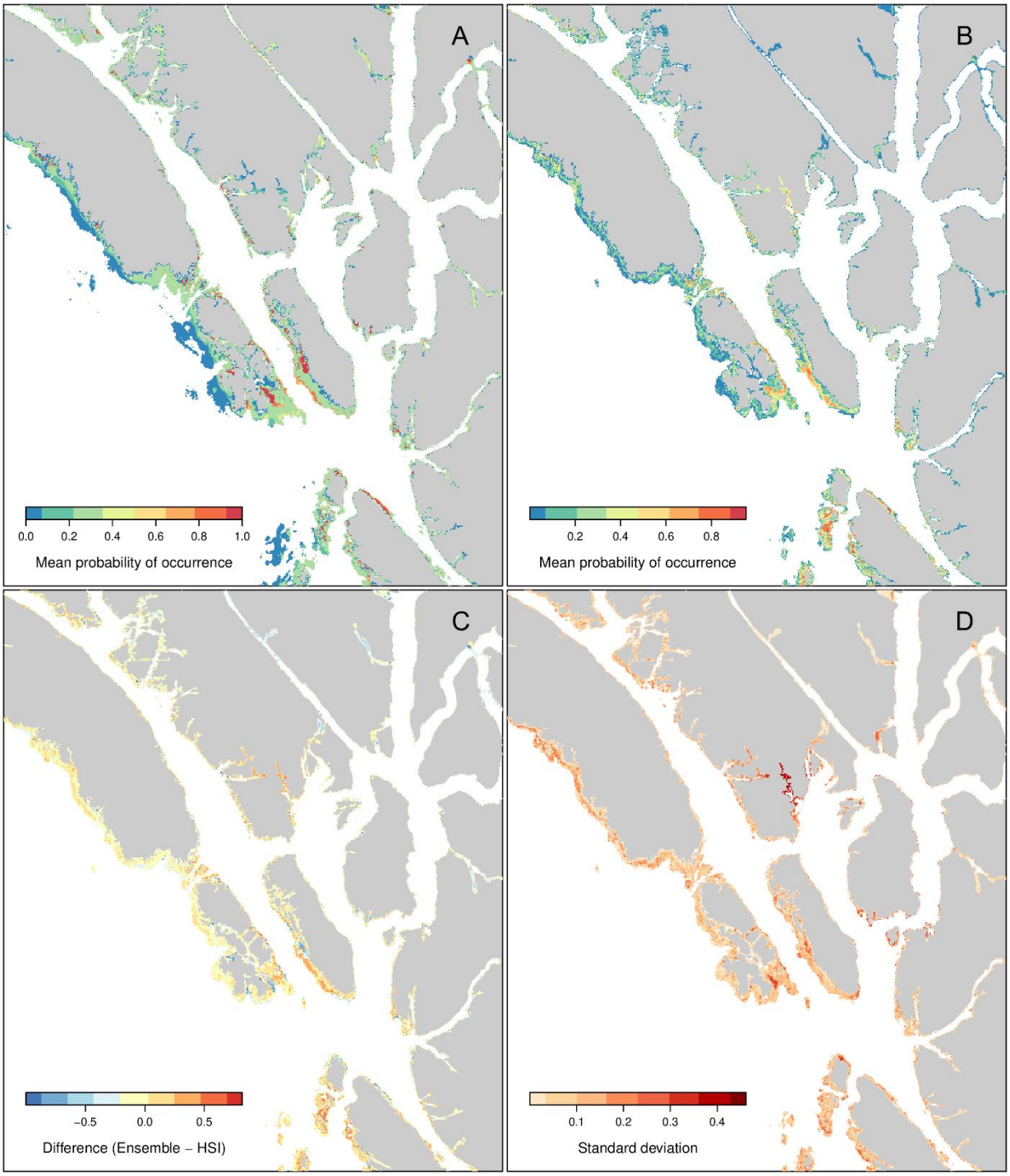

Figure A.2. Predictions of Geoduck distribution and the related uncertainty. Probability of occurrence predictions from A) the habitat suitability index model (HSI) and B) the ensemble model based on generalized linear and boosted regression tree models. Model uncertainty is represented by C) the difference between the HSI and the ensemble model predictions and D) the standard deviation across multiple ensemble model predictions. The area represented here is denoted in figure A.1 by the dashed box.

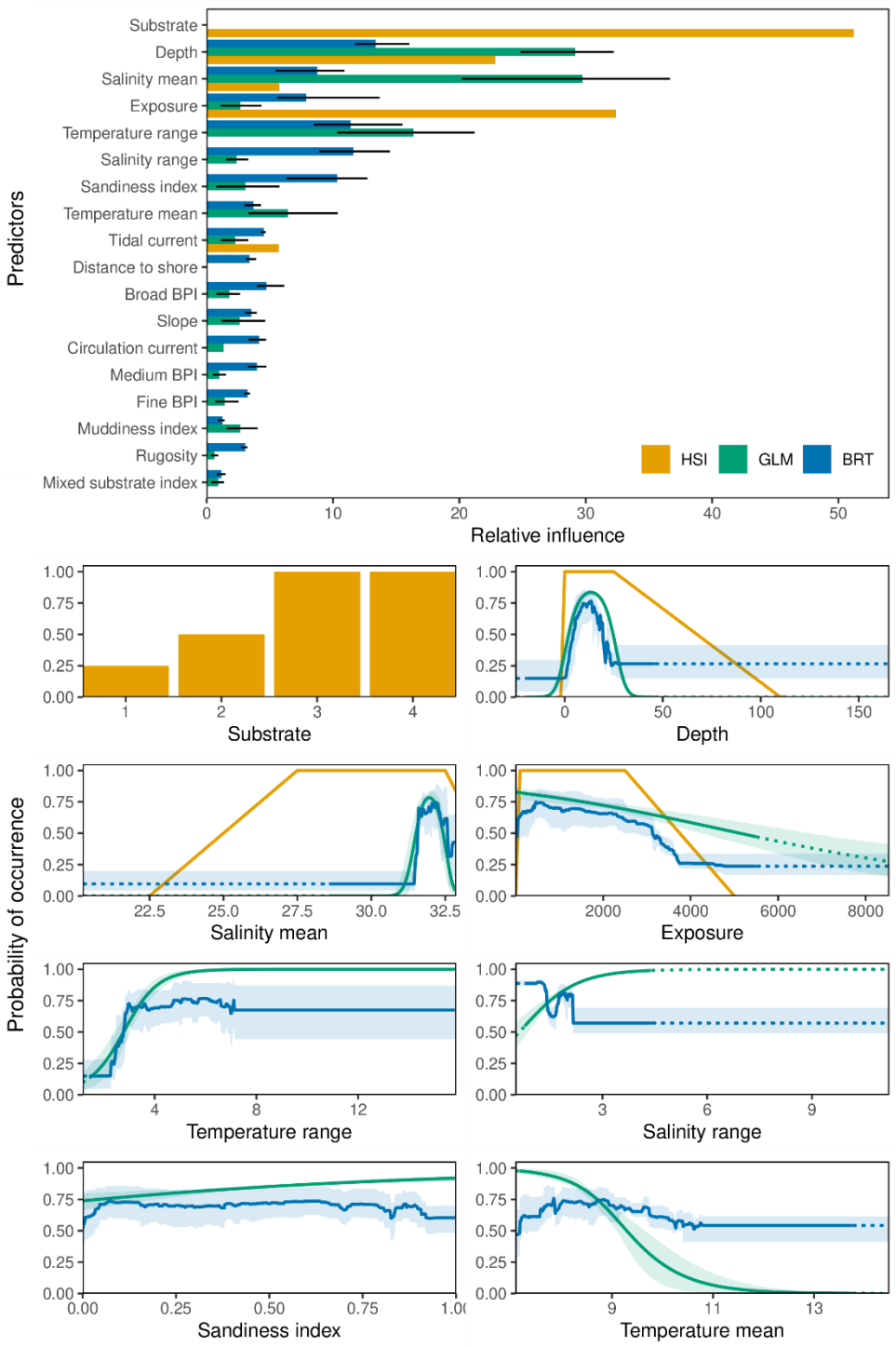

Figure A.3. Relative influence of predictors (top) and marginal effects (bottom, multi-panel) for the eight most influential environmental predictors from each of the HSI, GLM and BRT Geoduck models. For GLM and BRT models, the bars in the relative influence plots represent the mean and the lines show the minimum and maximum relative influence across the five-fold CV models. In the marginal effects plots, solid lines represent the mean marginal effects by method, and the shaded areas represents the minimum and maximum marginal effects across the five-fold CV models. Substrate was represented as a categorical variable for the HSI model and as a continuous sandiness index for GLM and BRT models.

Red Sea Urchin

Mesocentrotus franciscanus

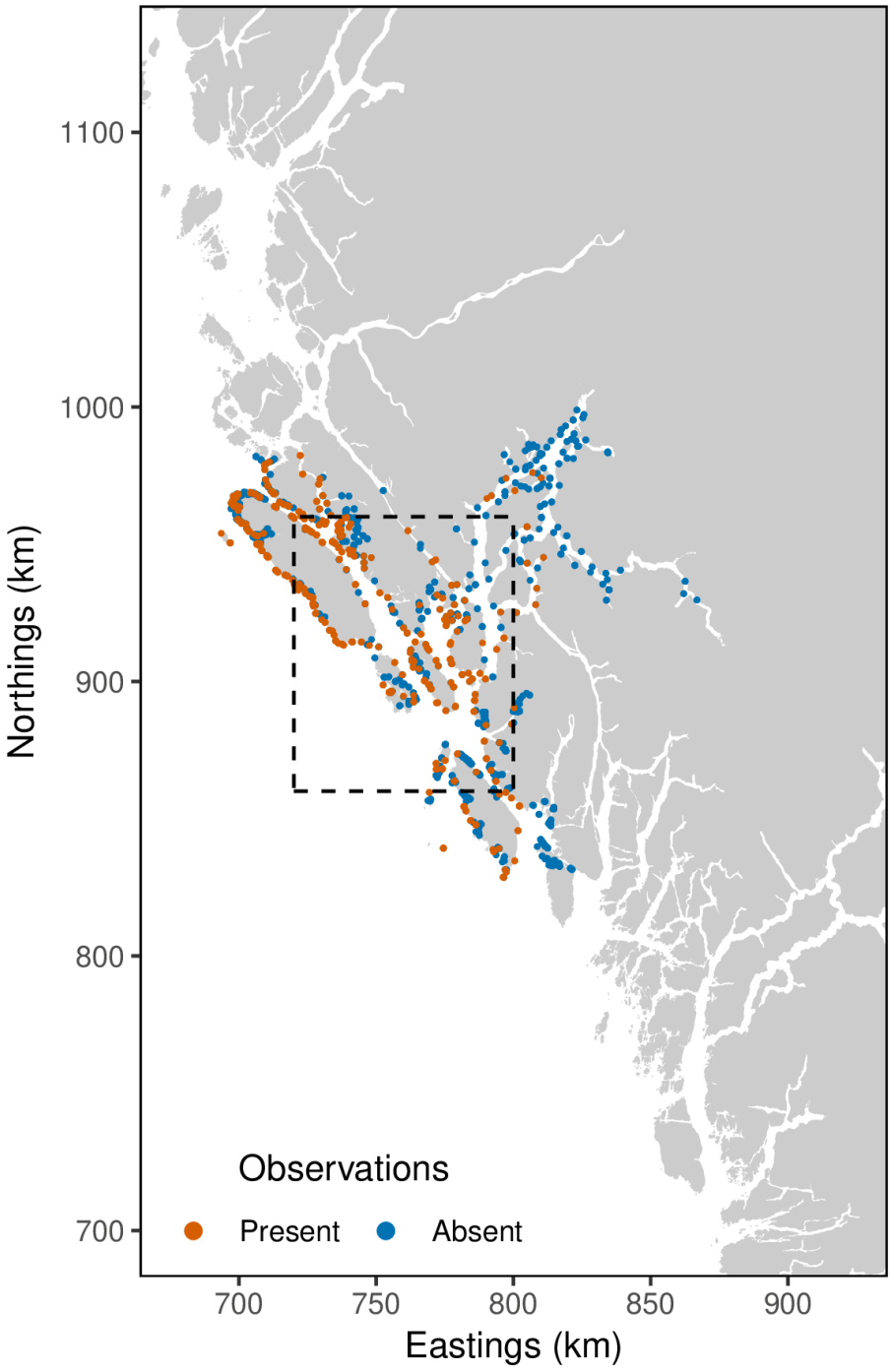

Figure A.4. Red Sea Urchin presence and absence observations within the nearshore study area. Dashed box shows the area displayed in figure A.5.

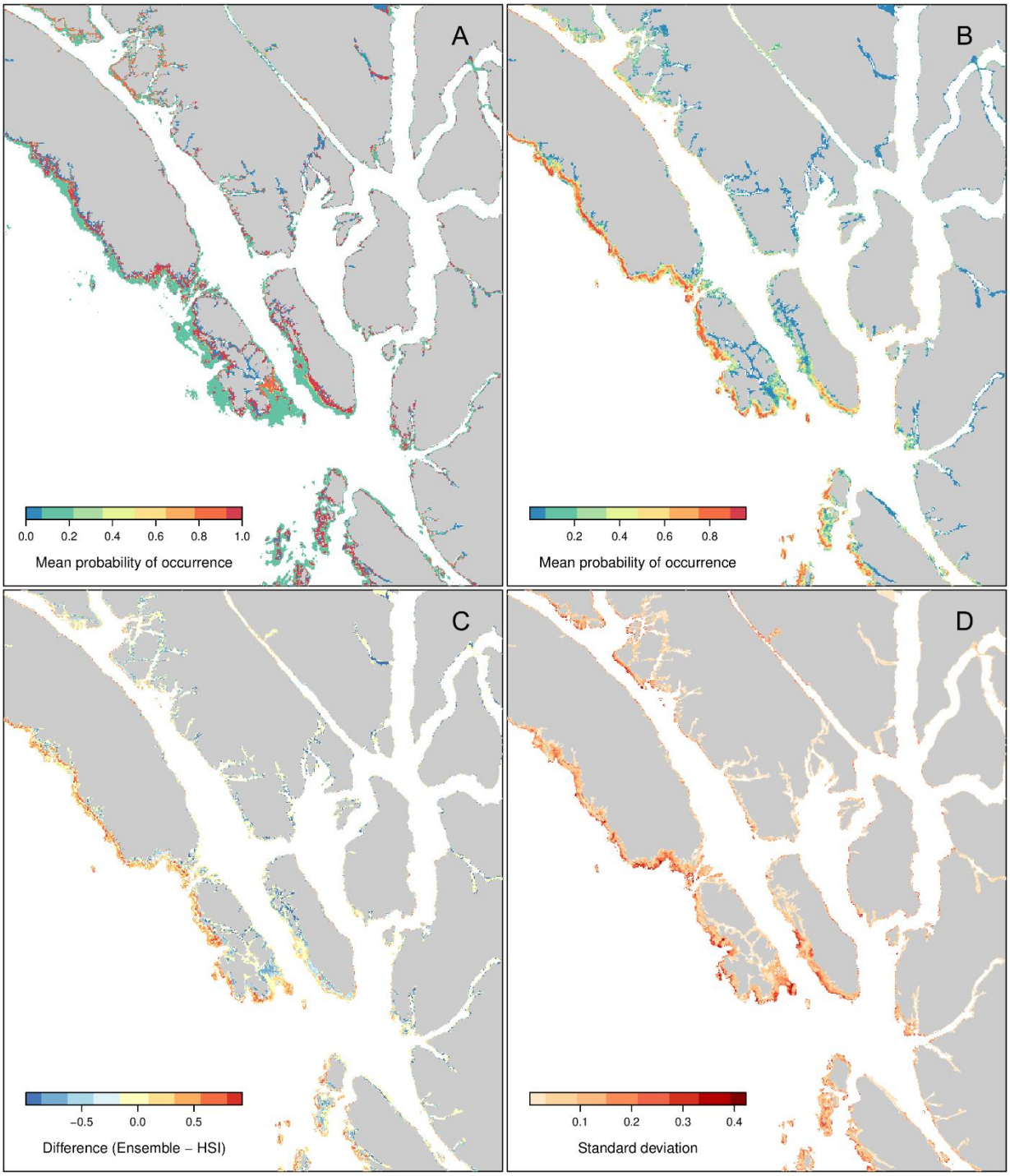

Figure A.5. Predictions of Red Sea Urchin distribution and the related uncertainty. Probability of occurrence predictions from A) the habitat suitability index model (HSI) and B) the ensemble model based on generalized linear and boosted regression tree models. Model uncertainty is represented by C) the difference between the HSI and the ensemble model predictions and D) the standard deviation across multiple ensemble model predictions. The area represented here is denoted in figure A.4 by the dashed box.

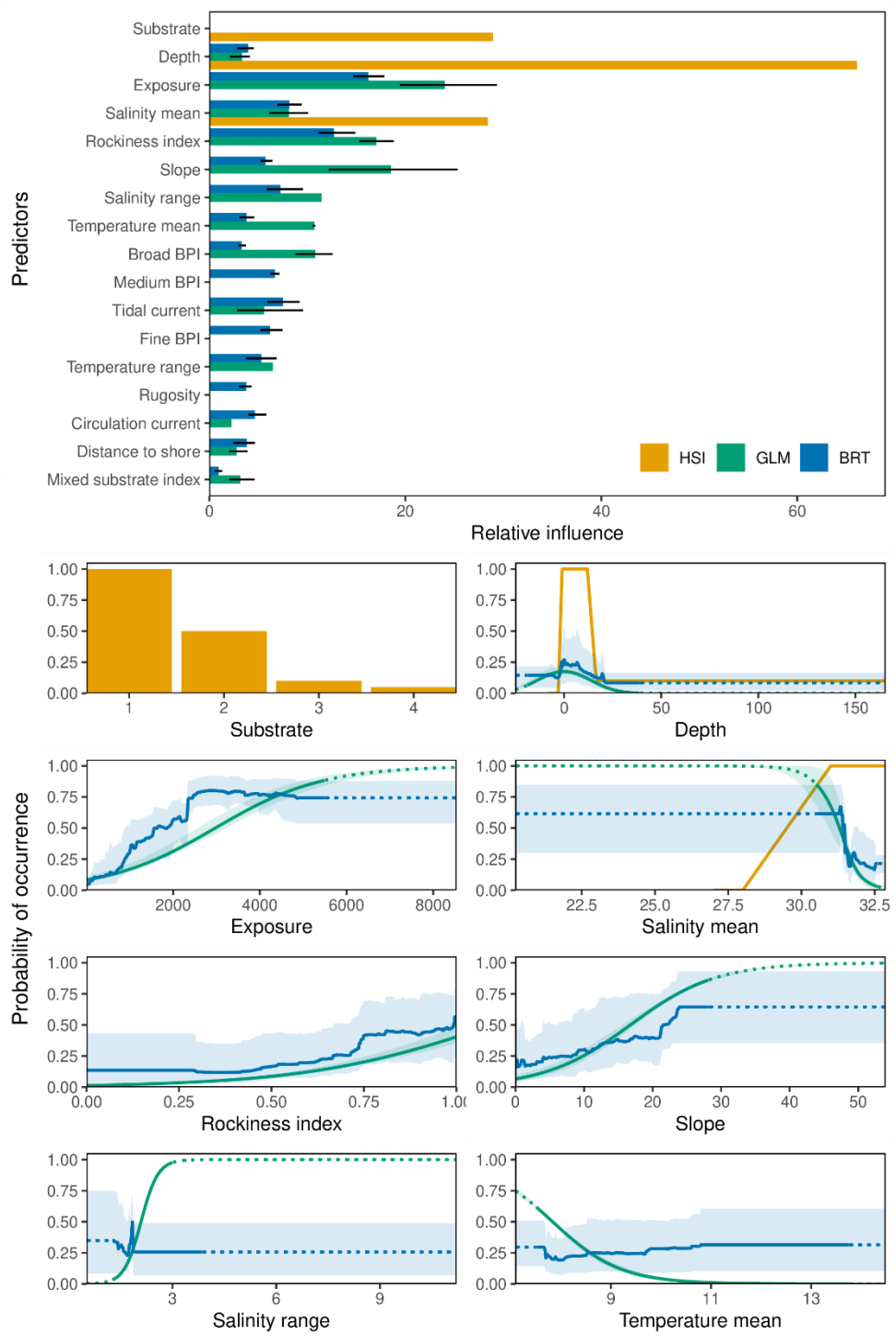

Figure A.6. Relative influence of predictors (top) and marginal effects (bottom, multi-panel) for the eight most influential environmental predictors from each of the HSI, GLM and BRT Red Sea Urchin models. For GLM and BRT models, the bars in the relative influence plots represent the mean and the lines show the minimum and maximum relative influence across the five-fold CV models. In the marginal effects plots, solid lines represent the mean marginal effects by method, and the shaded areas represents the minimum and maximum marginal effects across the five-fold CV models. Substrate was represented as a categorical variable for the HSI model and as a continuous rockiness index for GLM and BRT models.

Pterygophora Kelp

Pterygophora californica

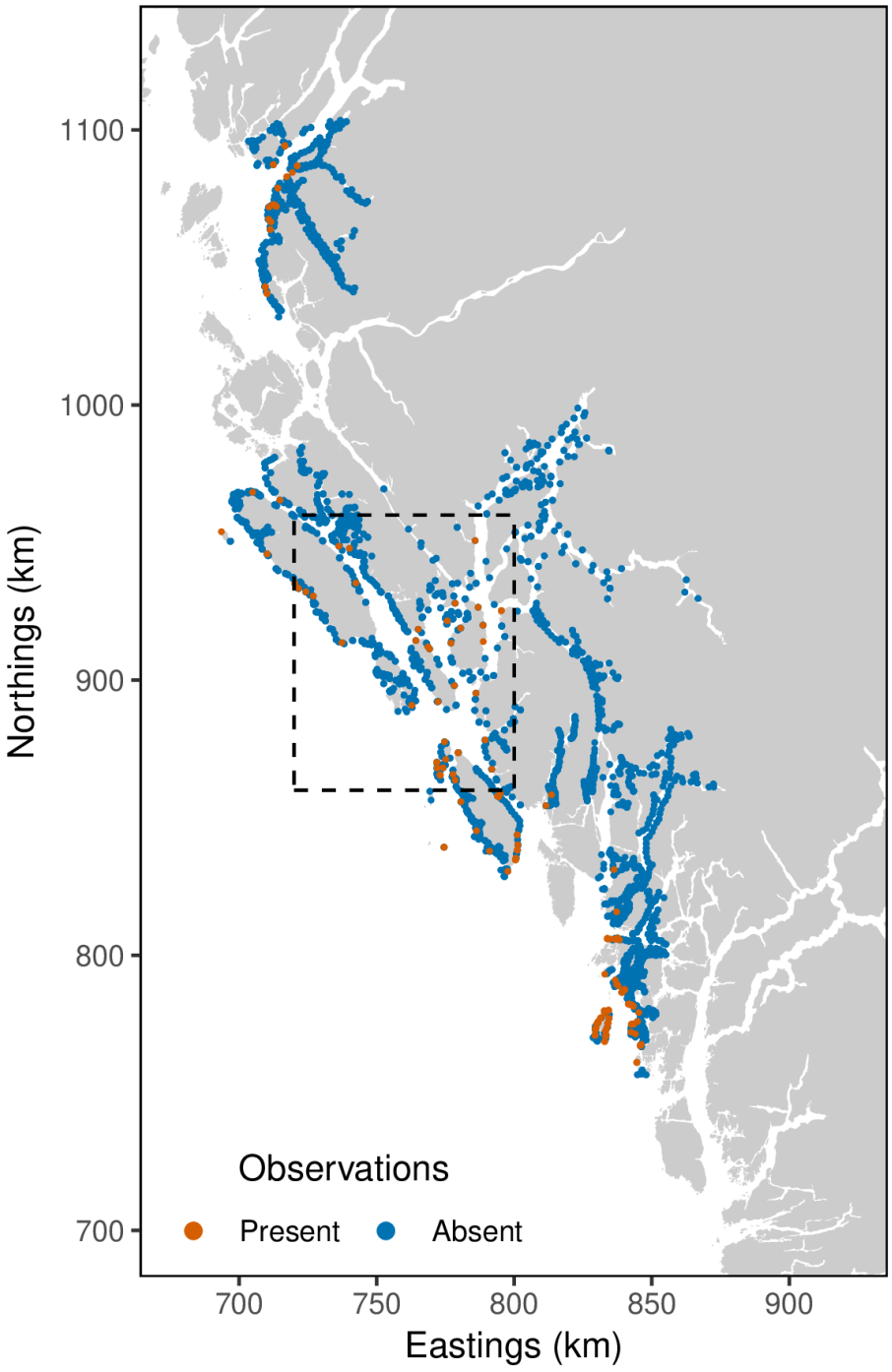

Figure A.7. Pterygophora Kelp presence and absence observations within the nearshore study area. Dashed box shows the area displayed in figure A.8.

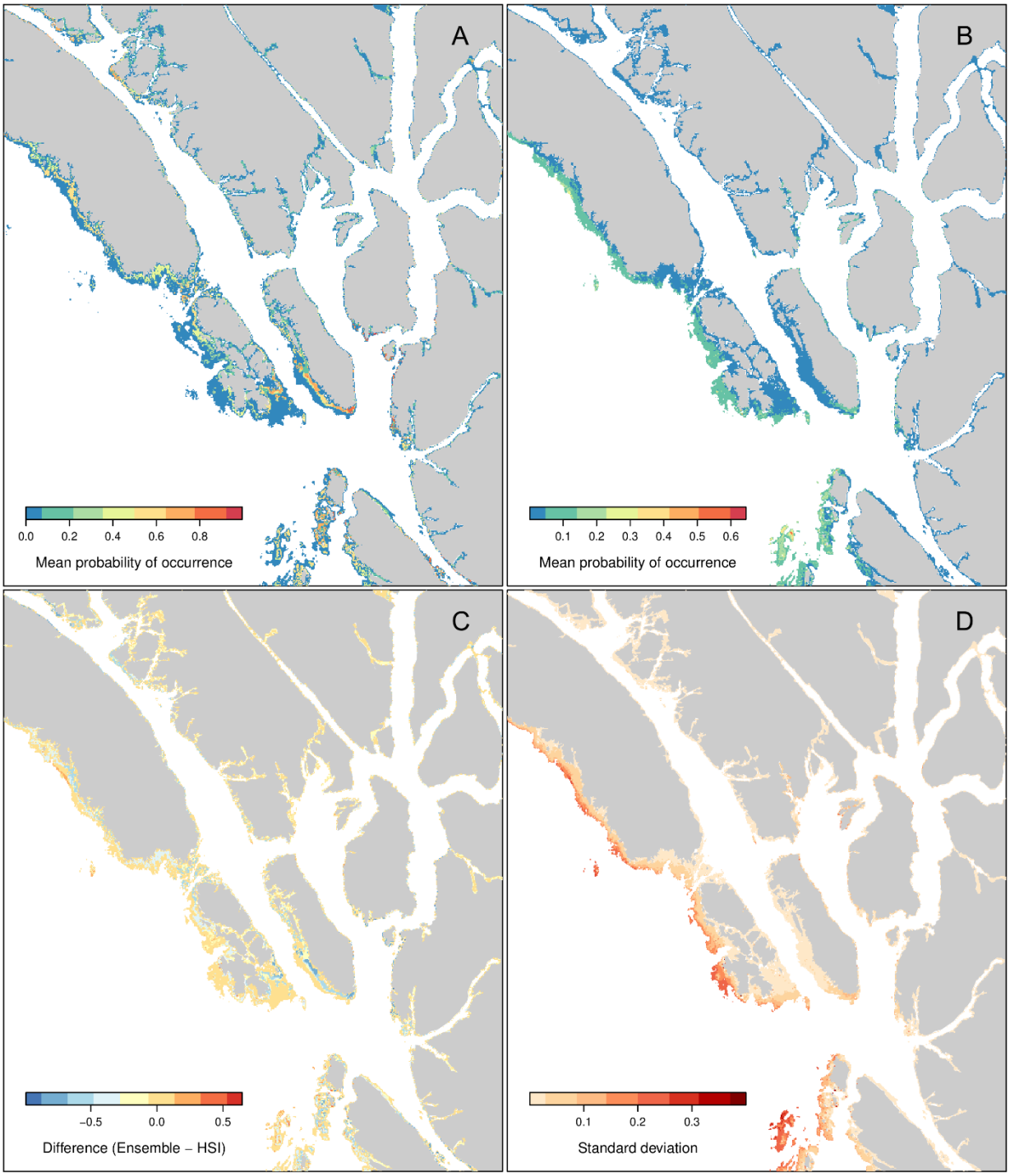

Figure A.8. Predictions of Pterygophora Kelp distribution and the related uncertainty. Probability of occurrence predictions from A) the habitat suitability index model (HSI) and B) the ensemble model based on generalized linear and boosted regression tree models. Model uncertainty is represented by C) the difference between the HSI and the ensemble model predictions and D) the standard deviation across multiple ensemble model predictions. The area represented here is denoted in figure A.7 by the dashed box.

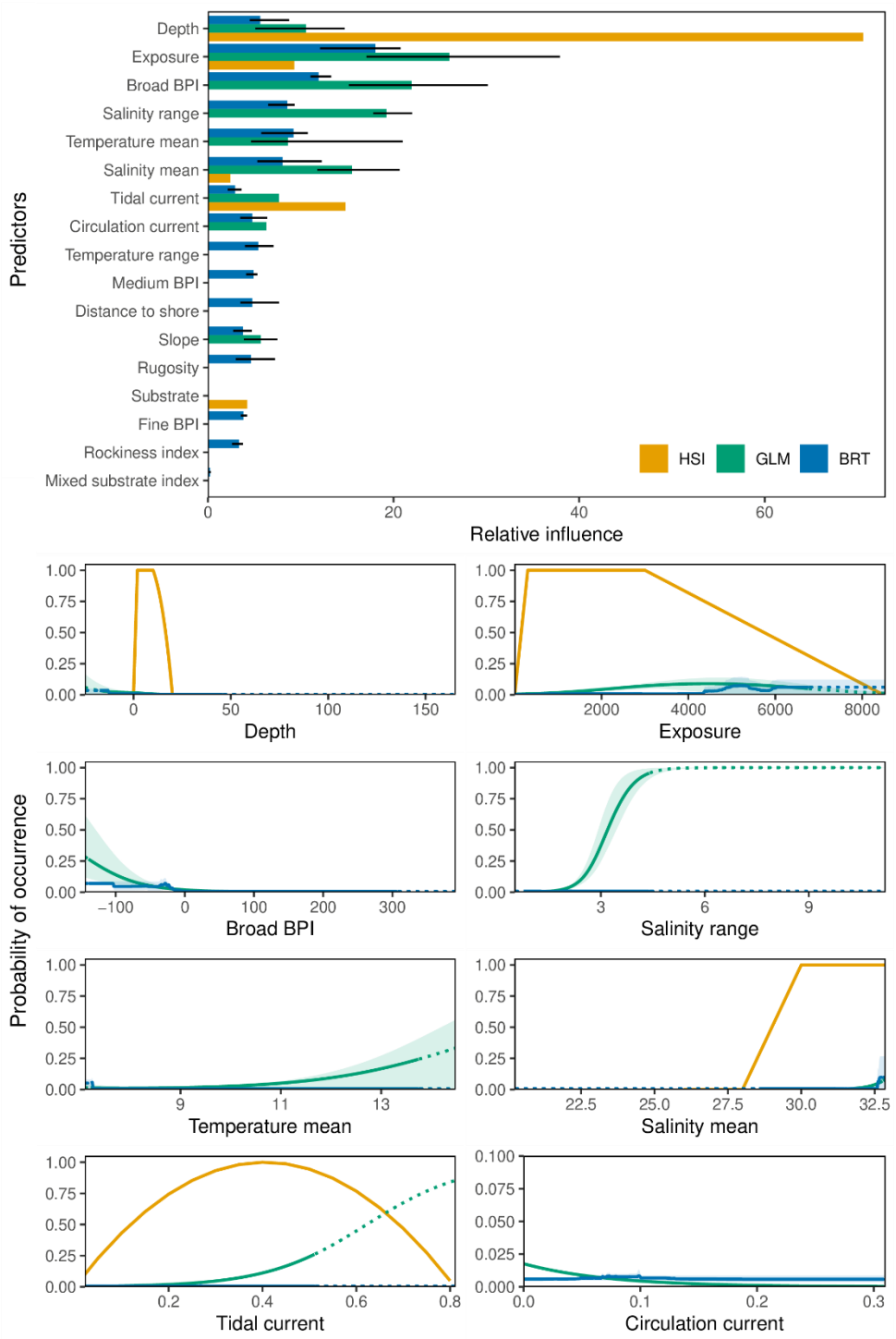

Figure A.9. Relative influence of predictors (top) and marginal effects (bottom, multi-panel) for the eight most influential environmental predictors from each of the HSI, GLM and BRT Pterygophora Kelp models. For GLM and BRT models, the bars in the relative influence plots represent the mean and the lines show the minimum and maximum relative influence across the five-fold CV models. In the marginal effects plots, solid lines represent the mean marginal effects by method, and the shaded areas represents the minimum and maximum marginal effects across the five-fold CV models. Substrate was represented as a categorical variable for the HSI model and as a continuous rockiness index for GLM and BRT models.

Eelgrass

Zostera spp.

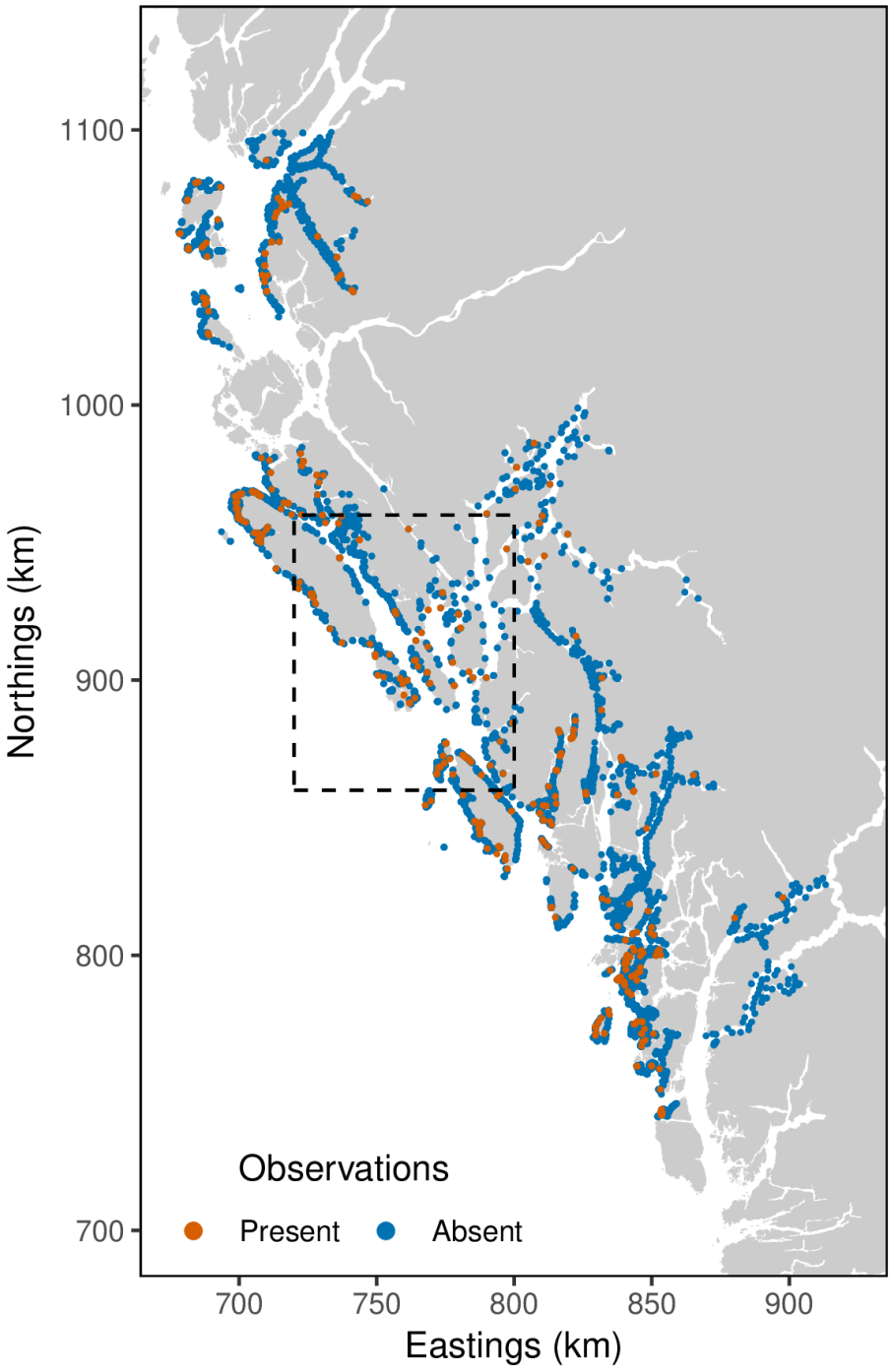

Figure A.10. Eelgrass presence and absence observations within the nearshore study area. Dashed box shows the area displayed in figure A.11.

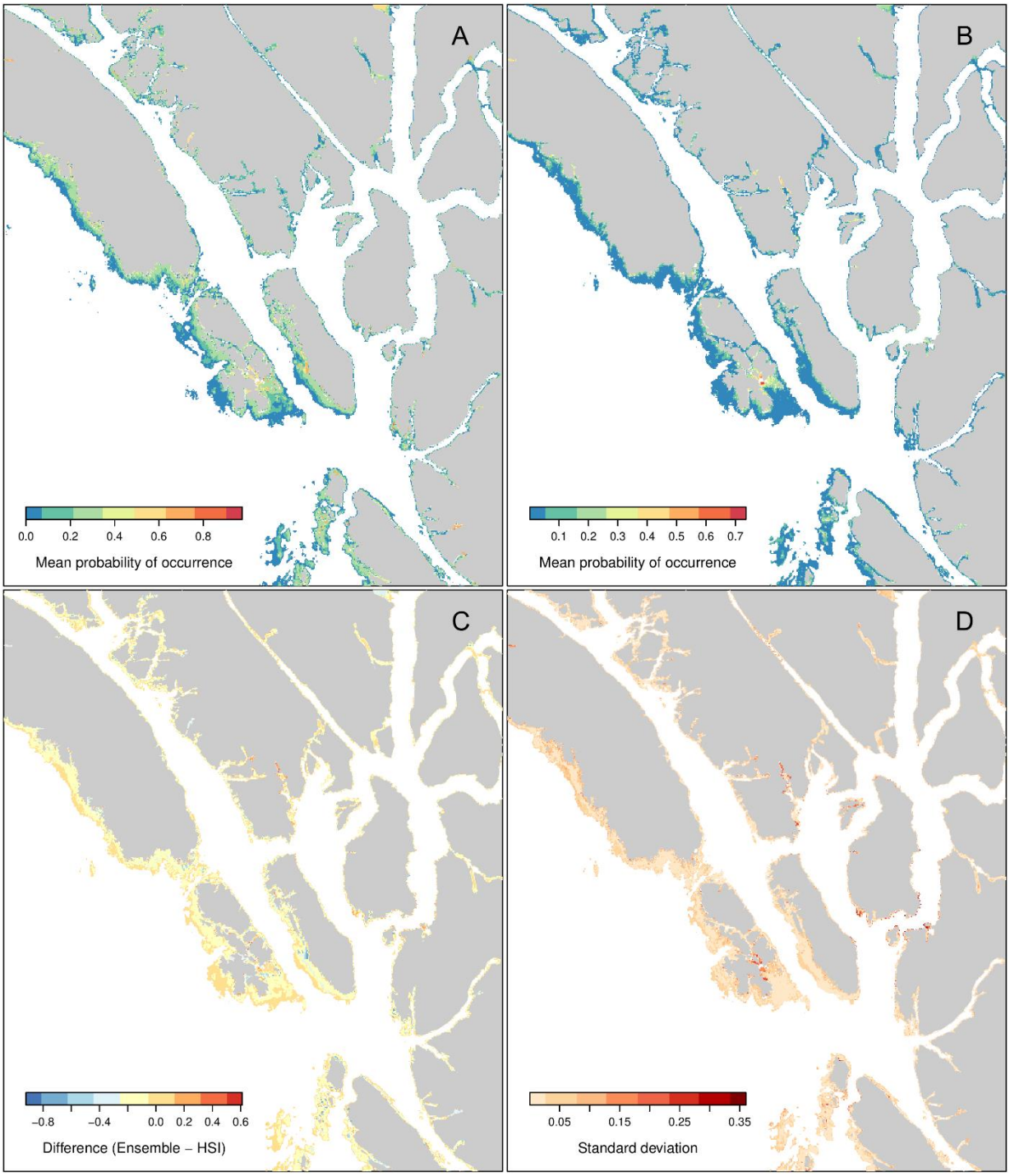

Figure A.11. Predictions of Eelgrass distribution and the related uncertainty. Probability of occurrence predictions from A) the habitat suitability index model (HSI) and B) the ensemble model based on generalized linear and boosted regression tree models. Model uncertainty is represented by C) the difference between the HSI and the ensemble model predictions and D) the standard deviation across multiple ensemble model predictions. The area represented here is denoted in figure A.10 by the dashed box.

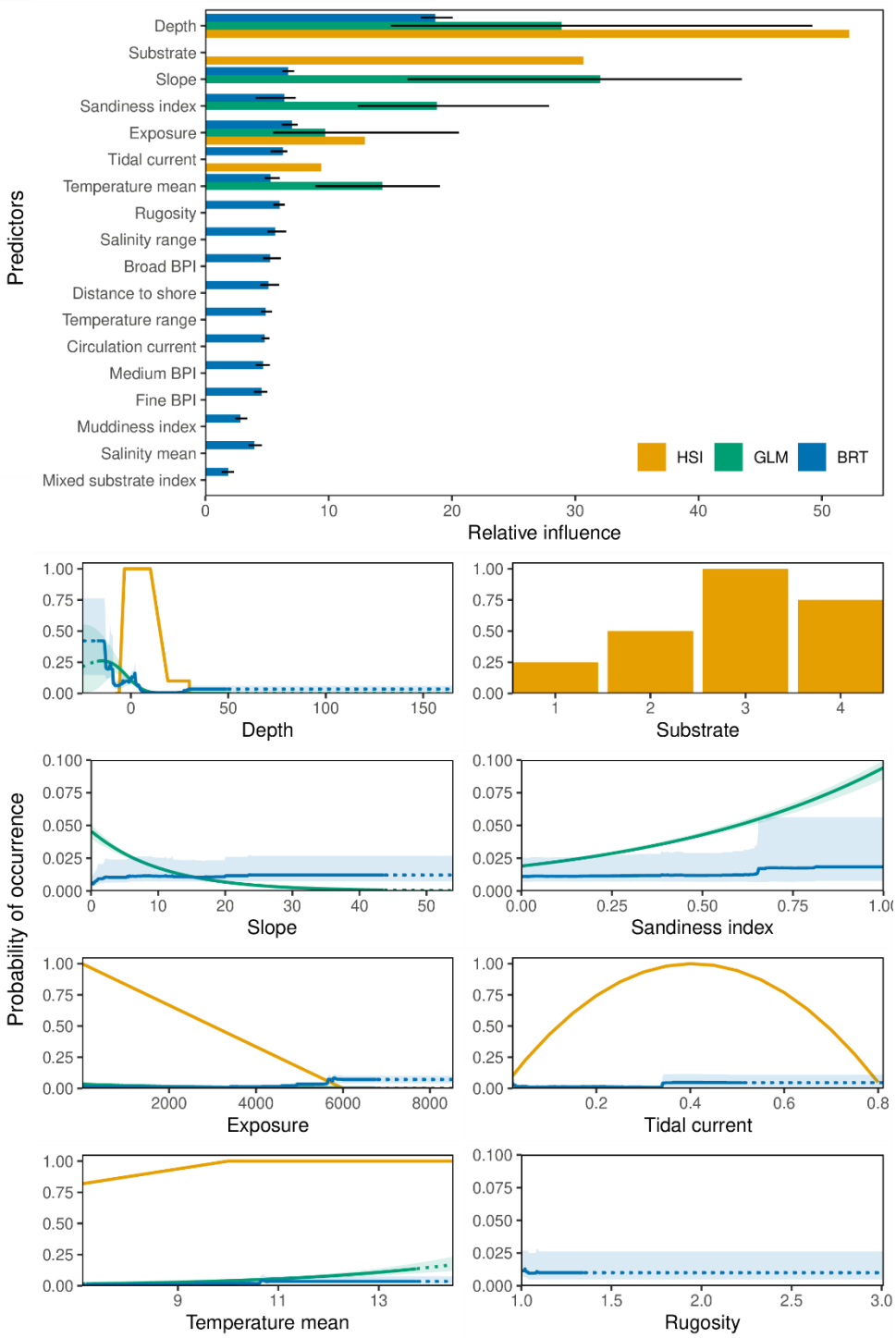

Figure A.12. Relative influence of predictors (top) and marginal effects (bottom, multi-panel) for the eight most influential environmental predictors from each of the HSI, GLM and BRT Eelgrass models. For GLM and BRT models, the bars in the relative influence plots represent the mean and the lines show the minimum and maximum relative influence across the five-fold CV models. In the marginal effects plots, solid lines represent the mean marginal effects by method, and the shaded areas represents the minimum and maximum marginal effects across the five-fold CV models. Substrate was represented as a categorical variable for the HSI model and as a continuous sandiness index for GLM and BRT models.

Dungeness Crab

Metacarcinus magister

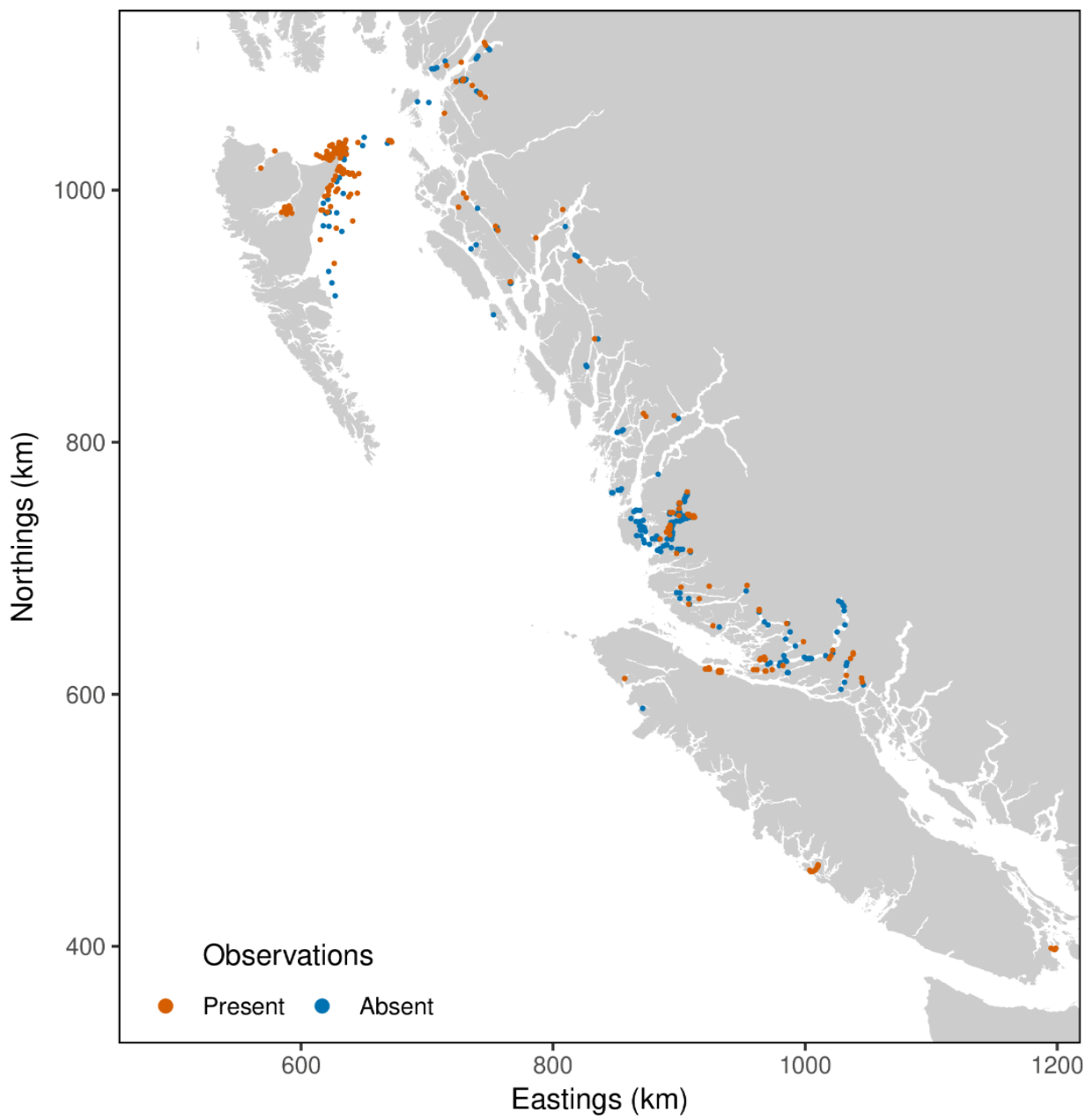

Figure A.13. Dungeness Crab presence and absence observations within the shelf study area.

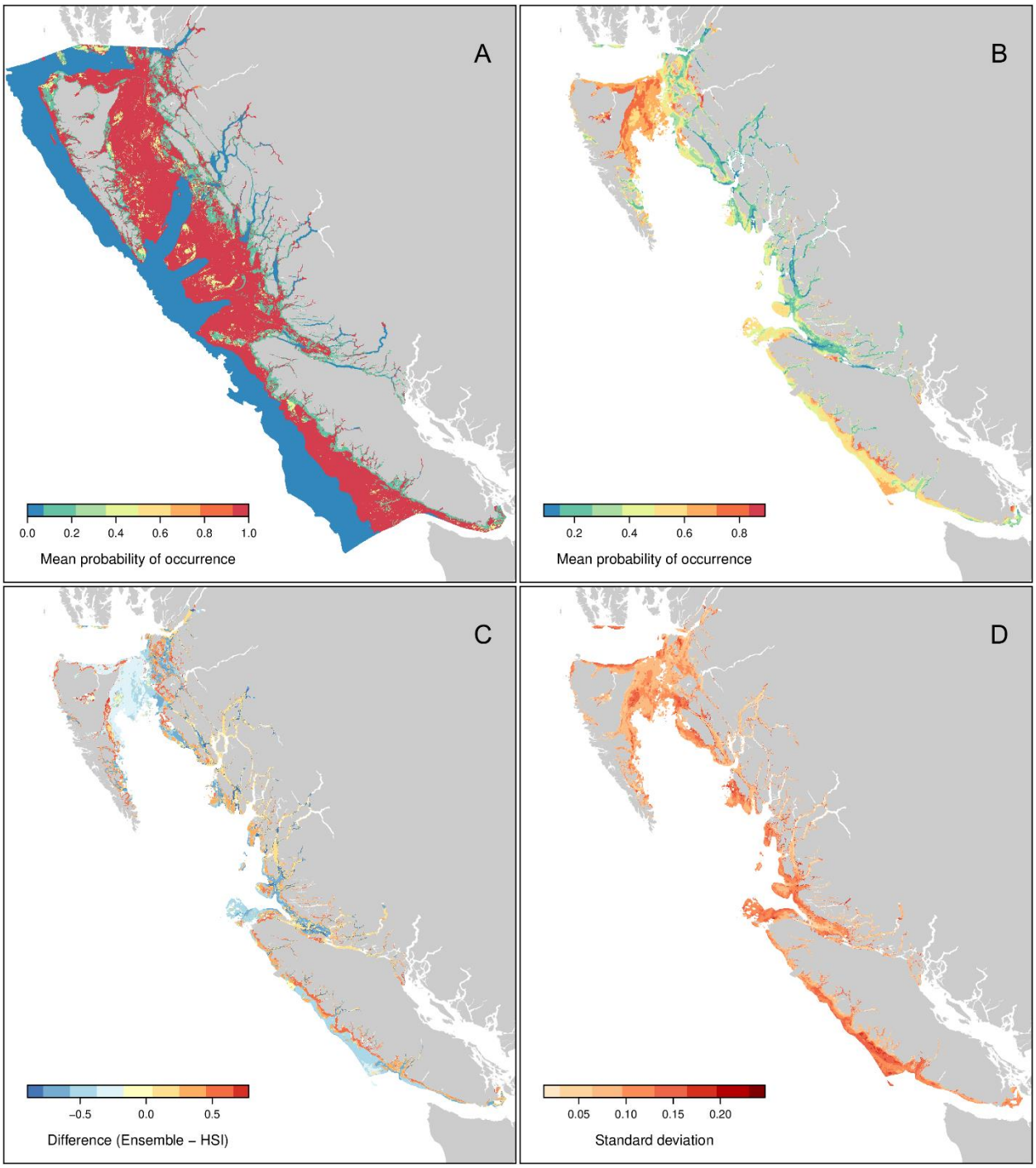

Figure A.14. Predictions of Dungeness Crab distribution and the related uncertainty. Probability of occurrence predictions from A) the habitat suitability index model (HSI) and B) the ensemble model based on generalized linear and boosted regression tree models. Model uncertainty is represented by C) the difference between the HSI and the ensemble model predictions and D) the standard deviation across multiple ensemble model predictions.

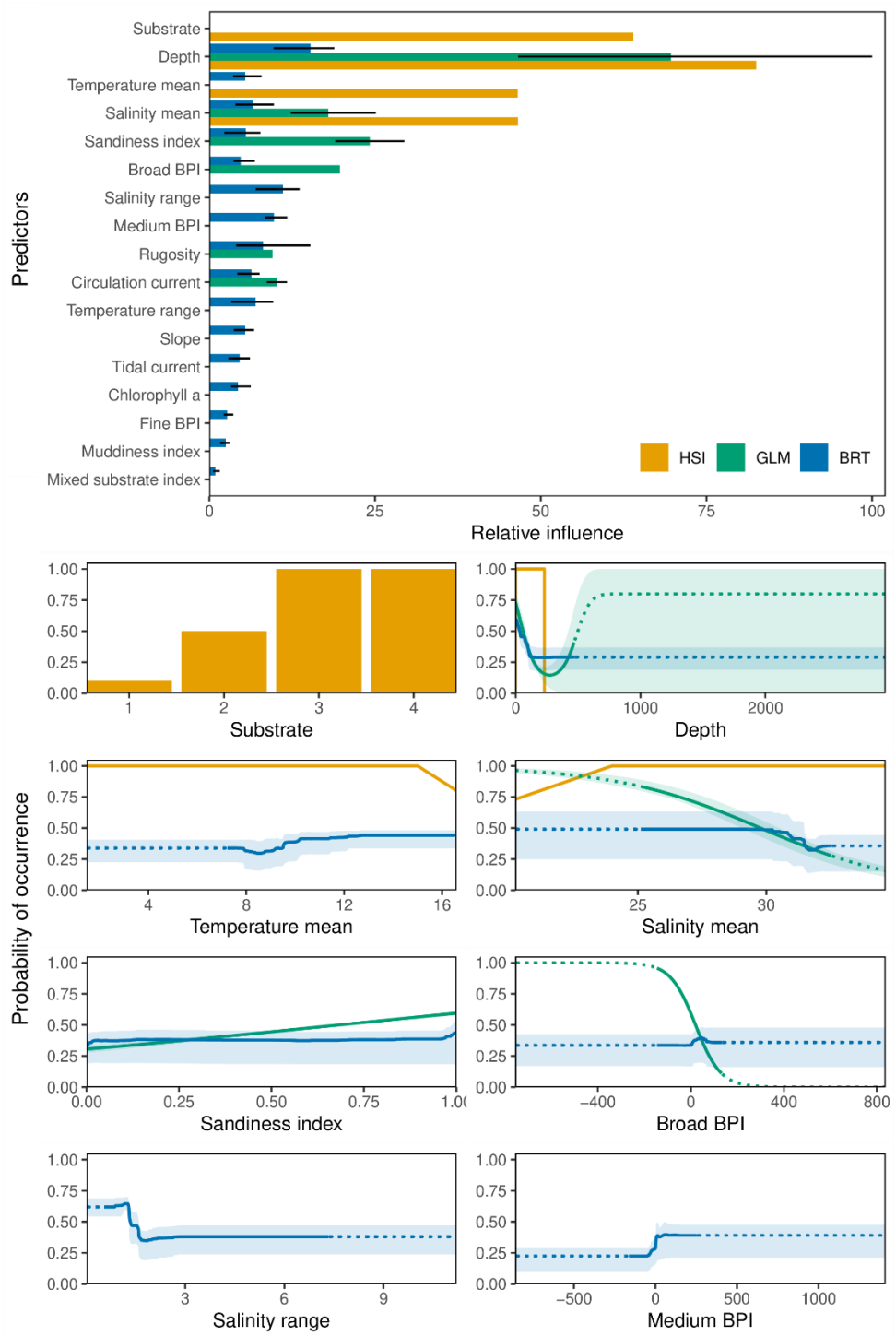

Figure A.15. Relative influence of predictors (top) and marginal effects (bottom, multi-panel) for the eight most influential environmental predictors from each of the HSI, GLM and BRT Dungeness Crab models. For GLM and BRT models, the bars in the relative influence plots represent the mean and the lines show the minimum and maximum relative influence across the five-fold CV models. In the marginal effects plots, solid lines represent the mean marginal effects by method, and the shaded areas represents the minimum and maximum marginal effects across the five-fold CV models. Substrate was represented as a categorical variable for the HSI model and as a continuous sandiness index for GLM and BRT models.

Quillback Rockfish

Sebastes maliger

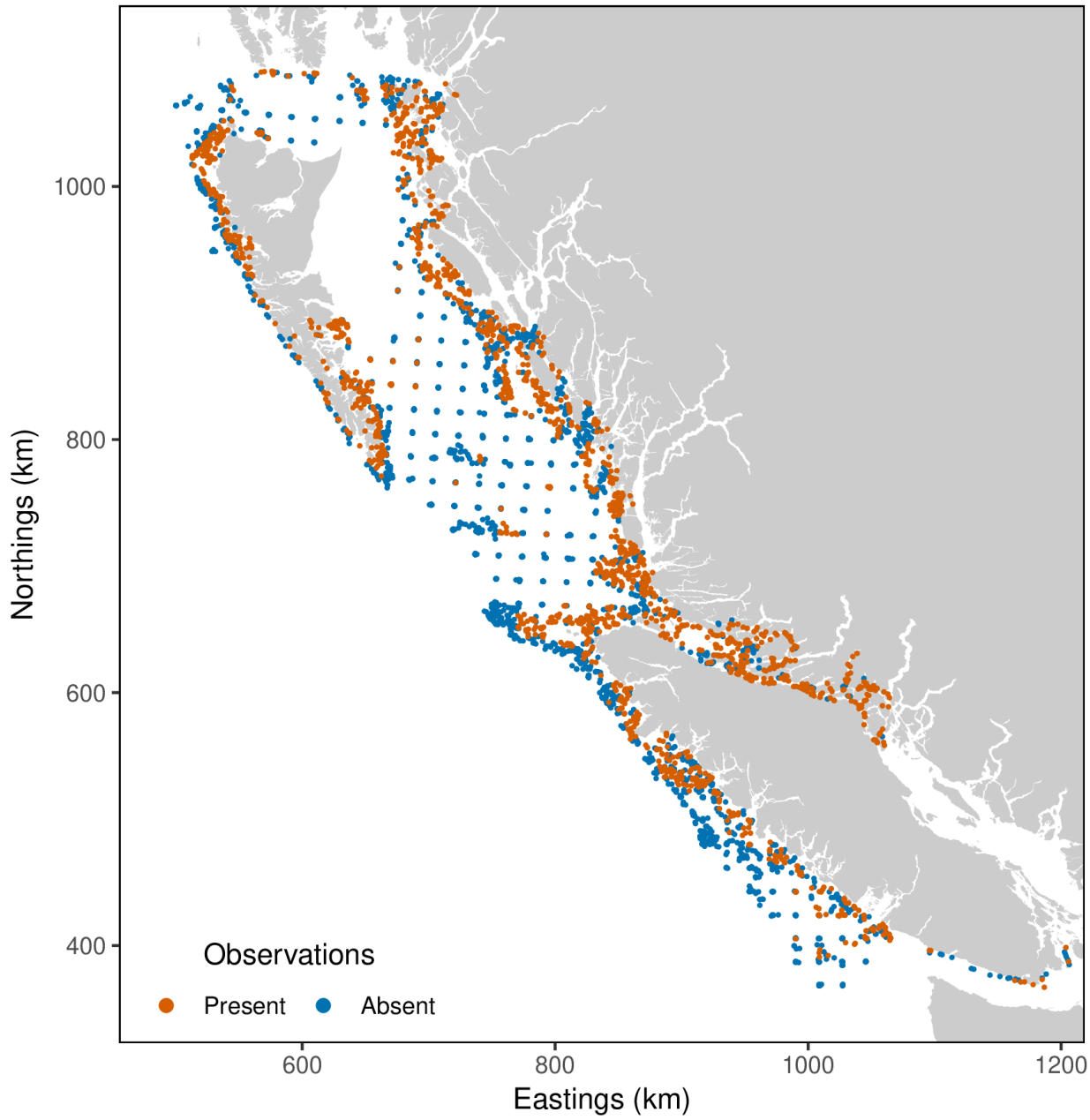

Figure A.16 Quillback Rockfish presence and absence observations within the shelf study area.

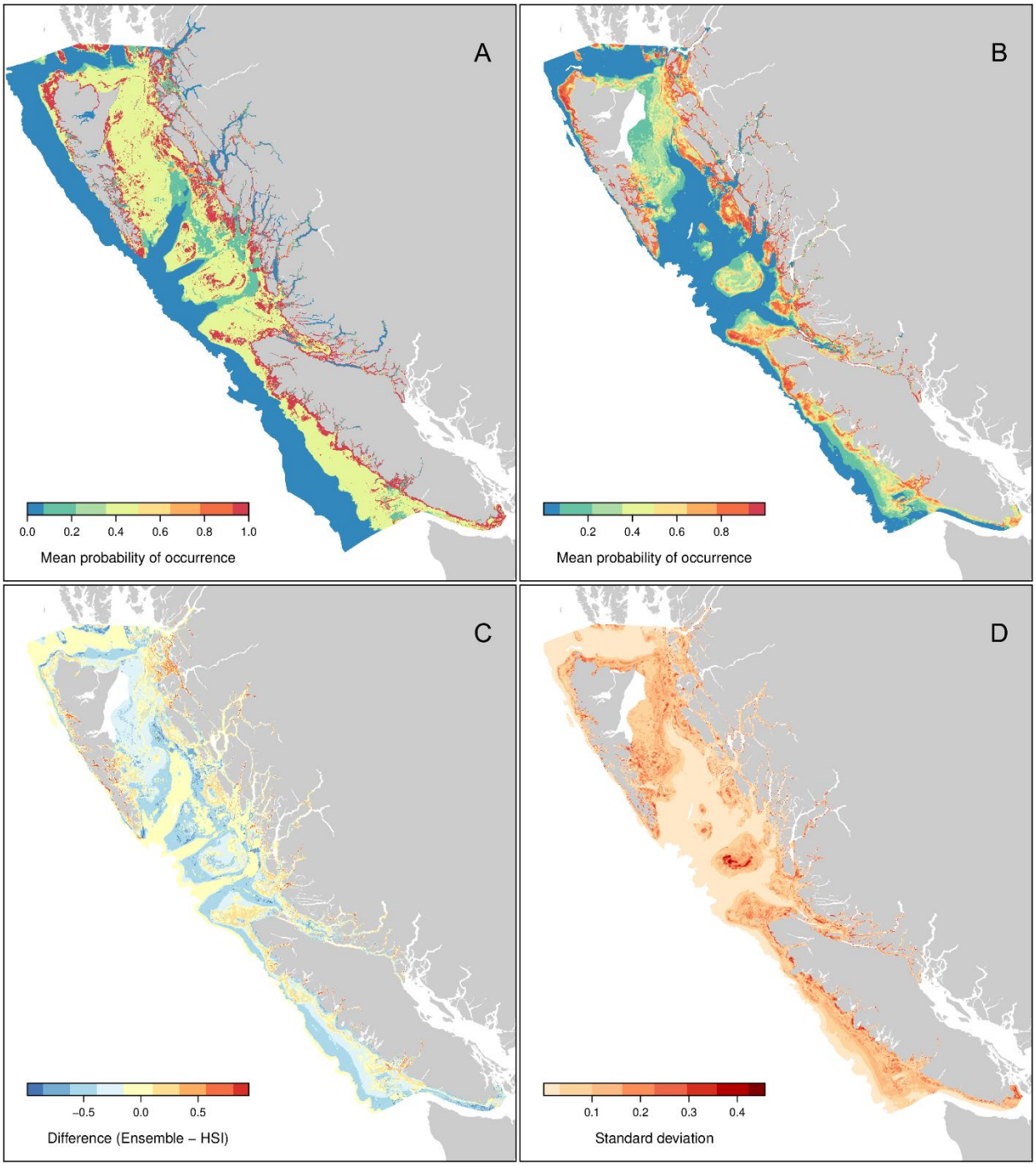

Figure A.17. Predictions of Quillback Rockfish distribution and the related uncertainty. Probability of occurrence predictions from A) the habitat suitability index model (HSI) and B) the ensemble model based on generalized linear and boosted regression tree models. Model uncertainty is represented by C) the difference between the HSI and the ensemble model predictions and D) the standard deviation across multiple ensemble model predictions.

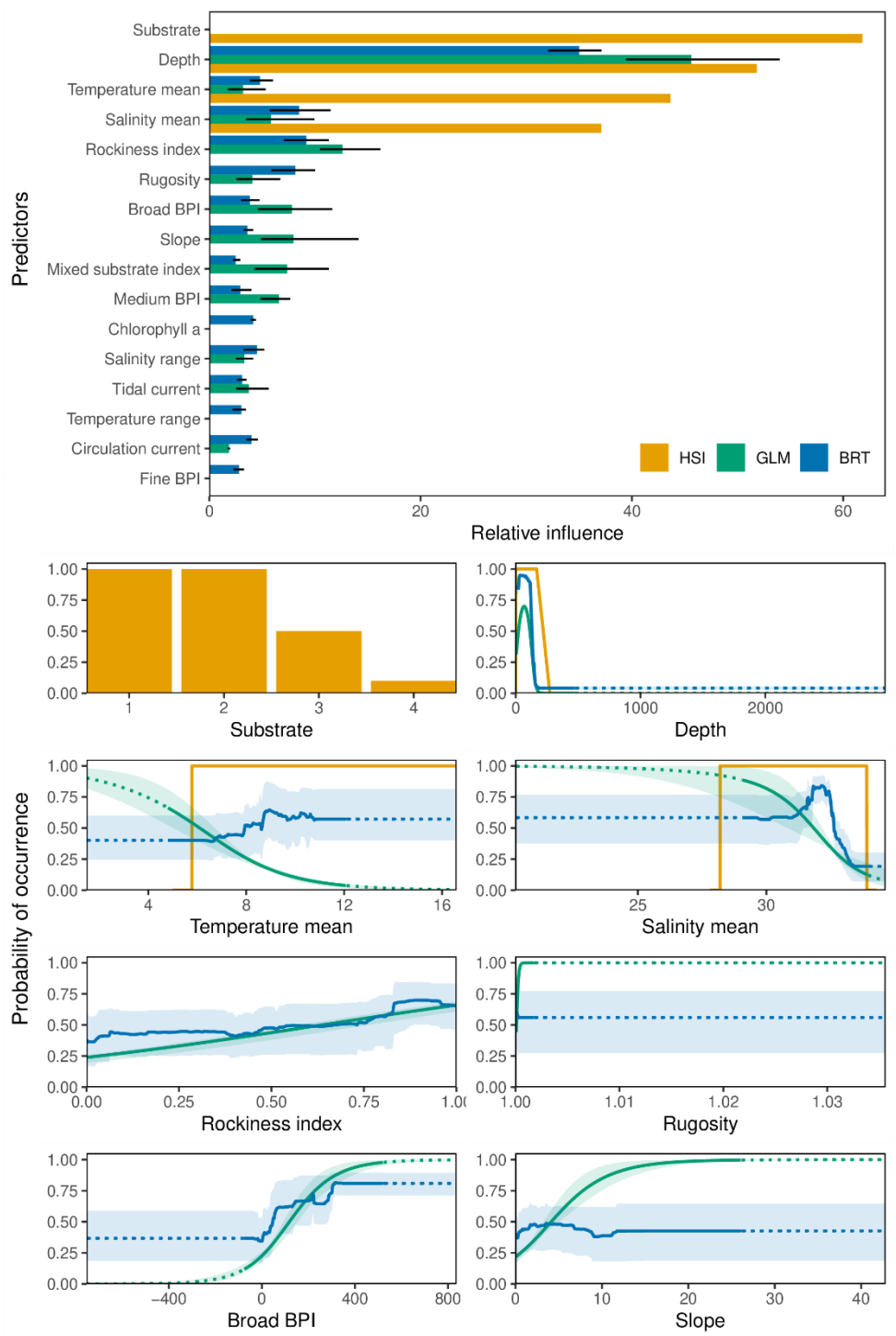

Figure A.18. Relative influence of predictors (top) and marginal effects (bottom, multi-panel) for the eight most influential environmental predictors from each of the HSI, GLM and BRT Quillback Rockfish models. For GLM and BRT models, the bars in the relative influence plots represent the mean and the lines show the minimum and maximum relative influence across the five-fold CV models. In the marginal effects plots, solid lines represent the mean marginal effects by method, and the shaded areas represents the minimum and maximum marginal effects across the five-fold CV models. Substrate was represented as a categorical variable for the HSI model and as a continuous rockiness index for GLM and BRT models.

Yelloweye Rockfish

Sebastes ruberrimus

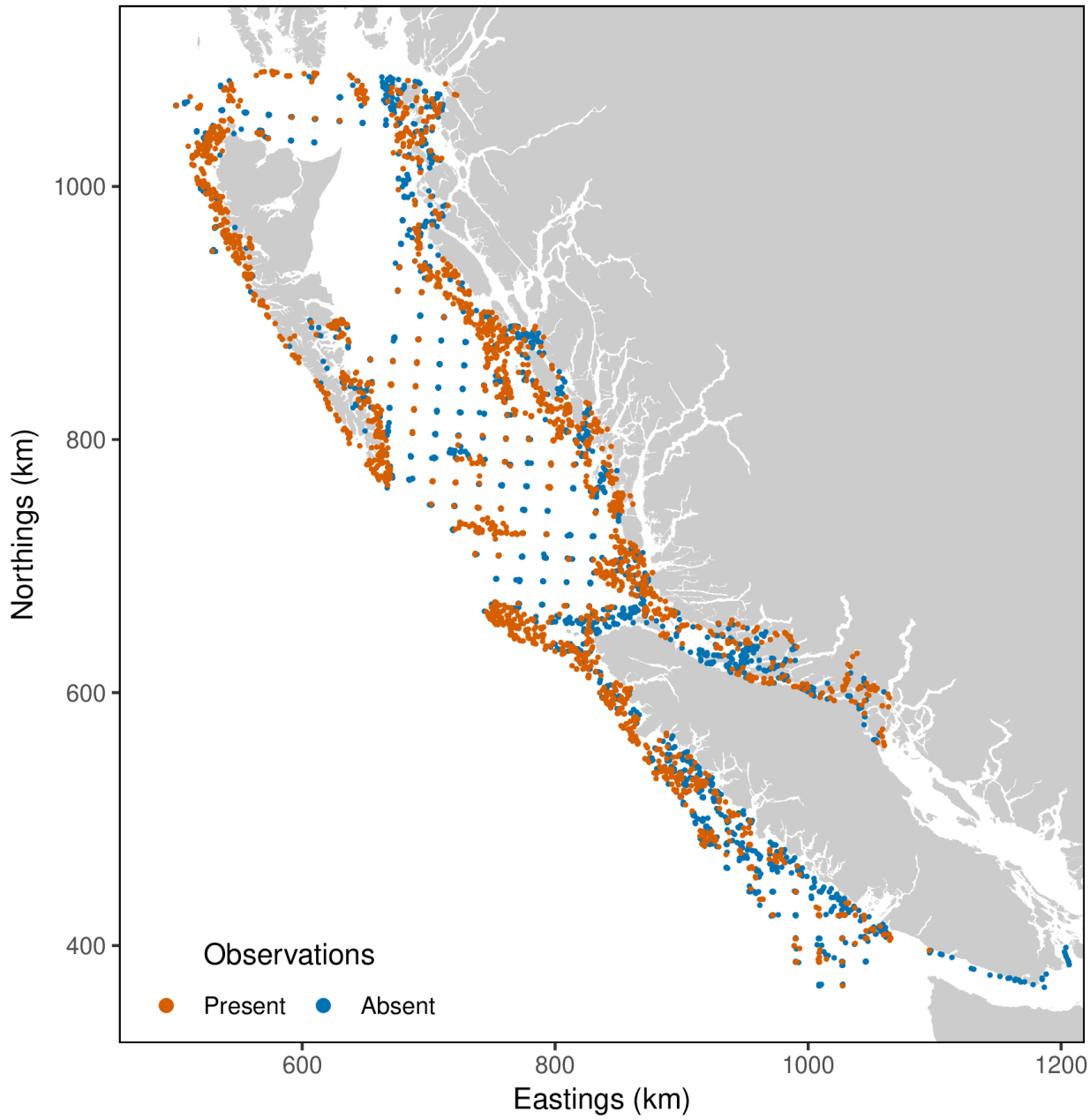

Figure A.19 Yelloweye Rockfish presence and absence observations within the shelf study area.

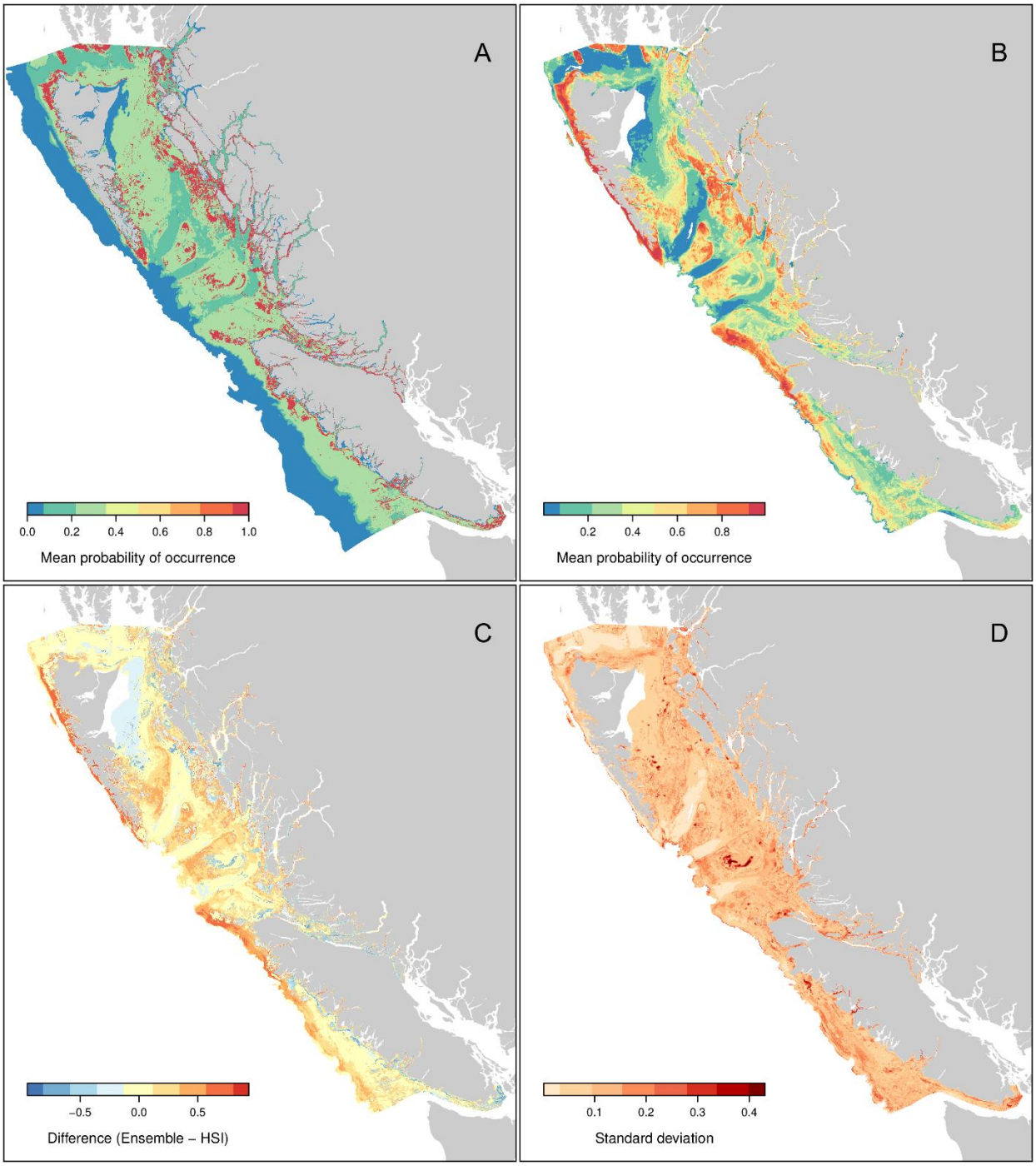

Figure A.20. Predictions of Yelloweye Rockfish distribution and the related uncertainty. Probability of occurrence predictions from A) the habitat suitability index model (HSI) and B) the ensemble model based on generalized linear and boosted regression tree models. Model uncertainty is represented by C) the difference between the HSI and the ensemble model predictions and D) the standard deviation across multiple ensemble model predictions.

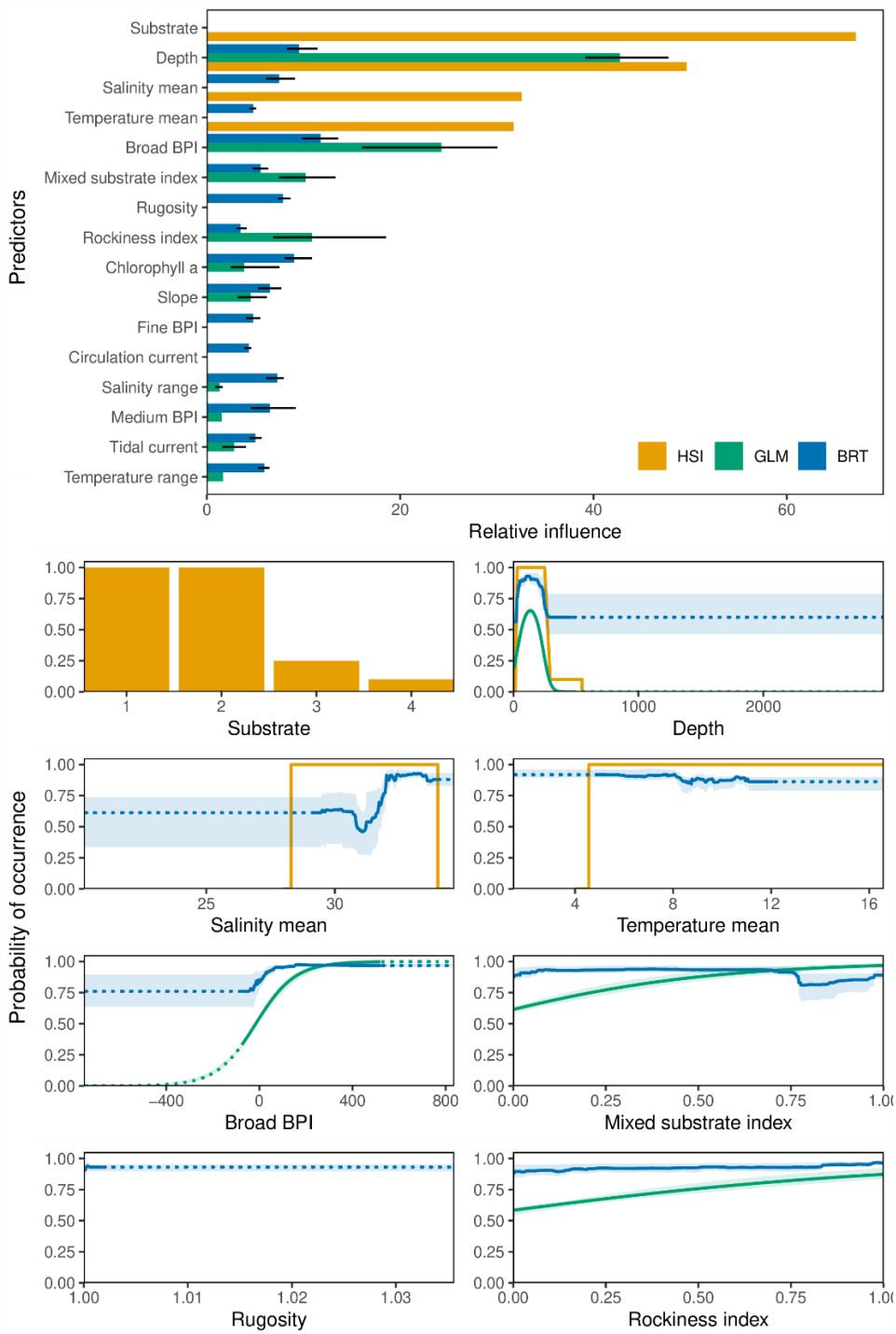

Figure A.21. Relative influence of predictors (top) and marginal effects (bottom, multi-panel) for the eight most influential environmental predictors from each of the HSI, GLM and BRT Yelloweye Rockfish models. For GLM and BRT models, the bars in the relative influence plots represent the mean and the lines show the minimum and maximum relative influence across the five-fold CV models. In the marginal effects plots, solid lines represent the mean marginal effects by method, and the shaded areas represents the minimum and maximum marginal effects across the five-fold CV models. Substrate was represented as a categorical variable for the HSI model and as a continuous rockiness index for GLM and BRT models.

Ochre Sea Star

Pisaster ochraceus

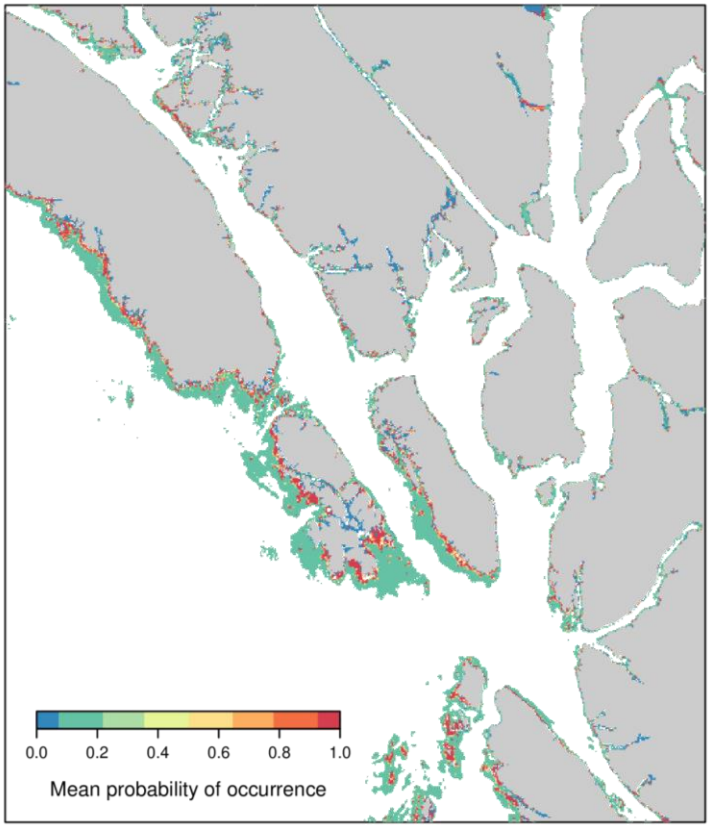

Figure A.22. Predictions of Ochre Sea Star distribution from a habitat suitability index model. The area represented here is denoted in figure A.1 by the dashed box.

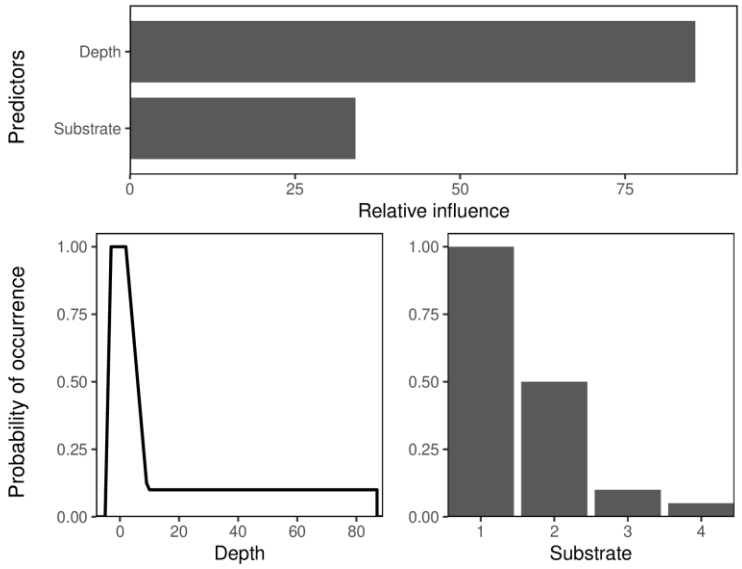

Figure A.23. Relative influence of predictors (top) and marginal effects (bottom, multi-panel) from the Ochre Sea Star habitat suitability index model.

Blue Mussel complex

Mytilus edulis, *M. trossulus*, and *M. galloprovincialis*

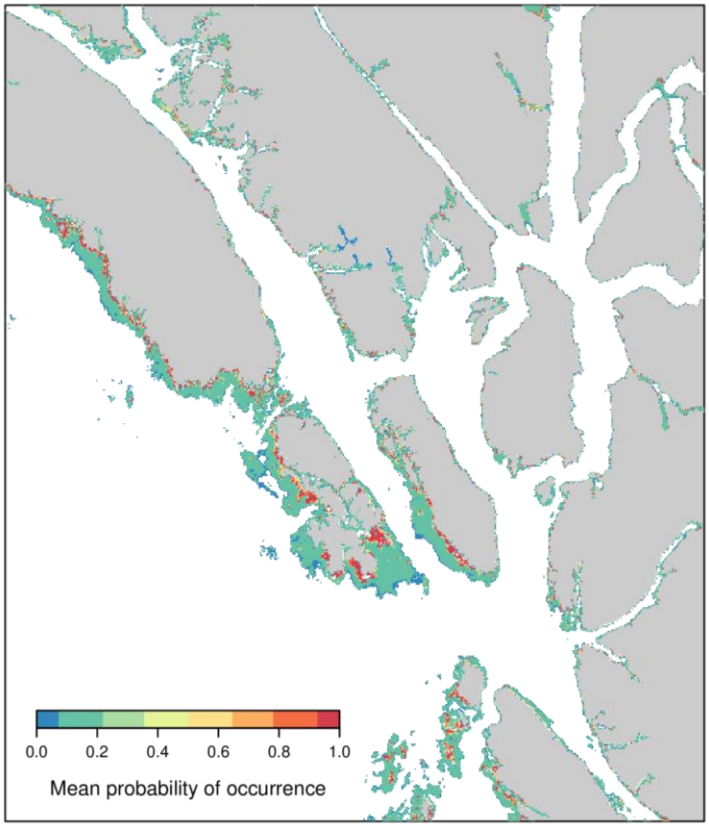

Figure A.24. Predictions of Blue Mussel distribution from a habitat suitability index model. The area represented here is denoted in figure A.1 by the dashed box.

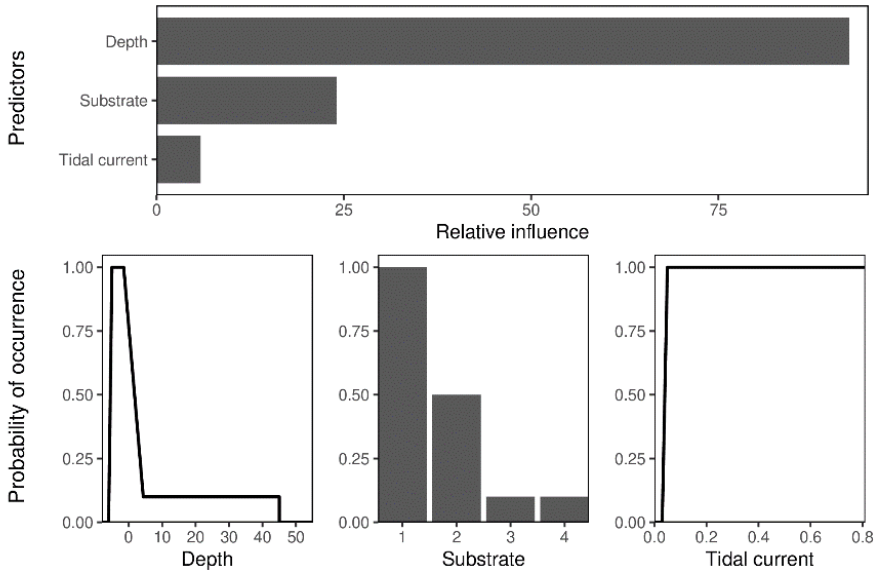

Figure A.25. Relative influence of predictors (top) and marginal effects (bottom, multi-panel) from the Blue Mussel habitat suitability index model.

Littleneck Clam

Leukoma staminea

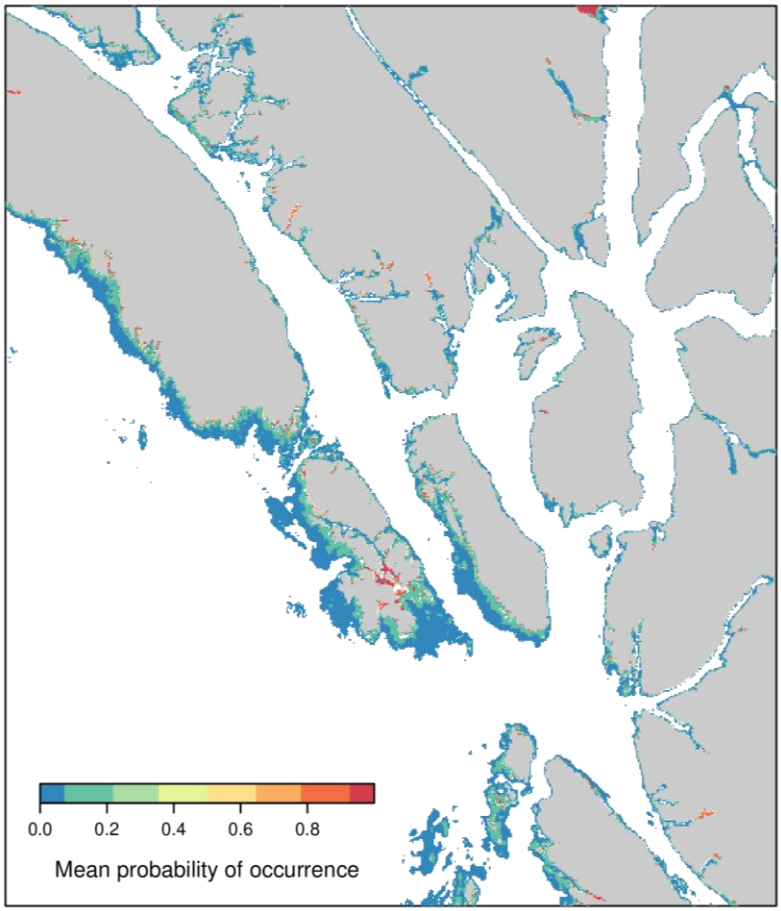

Figure A.26. Predictions of Littleneck Clam distribution from a habitat suitability index model. The area represented here is denoted in figure A.1 by the dashed box.

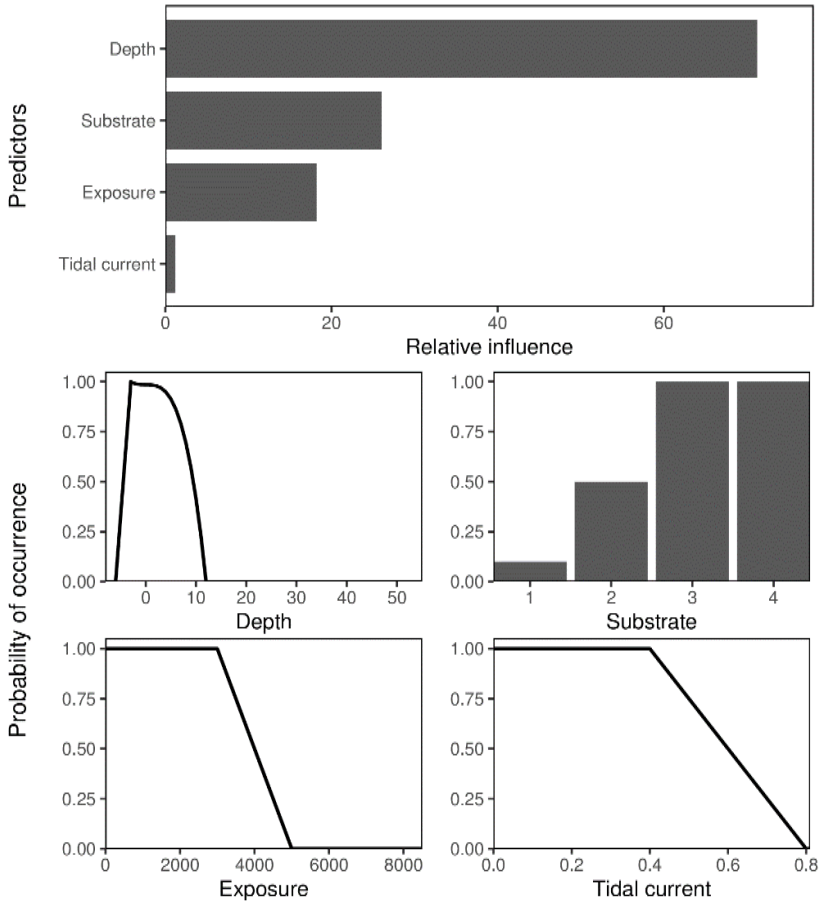

Figure A.27. Relative influence of predictors (top) and marginal effects (bottom, multi-panel) from the Littleneck Clam habitat suitability index model.

Orange Sea Pen

Ptilosarcus gurneyi

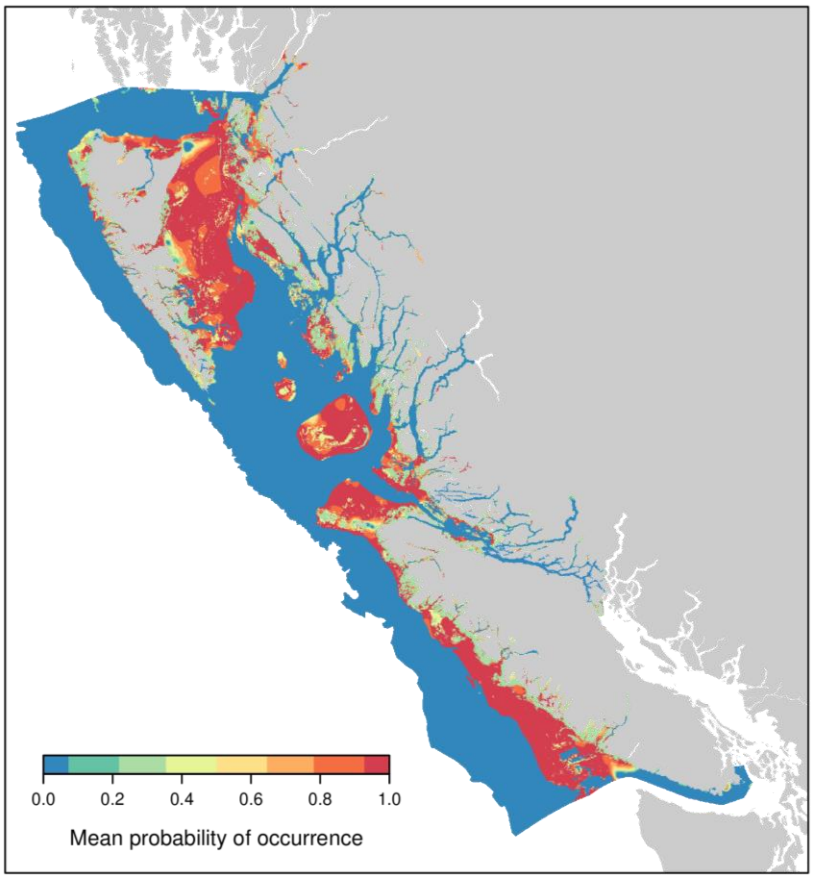

Figure A.28. Predictions of Orange Sea Pen distribution from a habitat suitability index model.

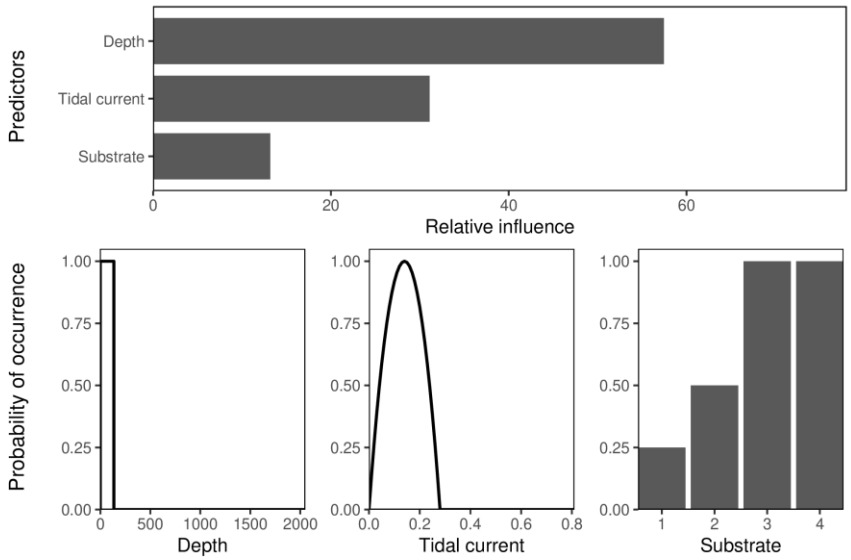

Figure A.29. Relative influence of predictors (top) and marginal effects (bottom, multi-panel) from the Orange Sea Pen habitat suitability index model.

APPENDIX B. HABITAT SUITABILITY INDEX DESCRIPTIONS

Habitat suitability index (HSI) models are envelope models that rely on expert opinion to derive hypotheses to relate habitat suitability to abiotic factors. The methods for building HSI models are detailed in Section 4.2.5. Here we outline the relevant ecology and assumptions made describing the environmental (abiotic) predictor relationships considered for each species. Knowledge related specifically to British Columbia (BC) was treated preferentially. In its absence, knowledge from the broader eastern North Pacific was used. All predictors are assumed to be limiting, making the suitability index at any particular location equal to the lowest occurring predictor value.

Predictor relationships were developed specifically for the following environmental predictor layers (described in Section 4.2.2) where appropriate: substrate (1 = rock, 2 = mixed, 3 = sand, 4 = mud); depth (m); mean summer bottom salinity (PSU); mean summer bottom temperature (°C); mean summer tidal current speed ($m \cdot s^{-1}$); and exposure (100s of km). The relationship between environmental variables and habitat suitability can change based on the temporal and spatial scales of those predictors. Thus, when building expert-derived habitat envelopes, fine scale ecological processes are not always captured. As higher resolution predictor layers become available, the relationships described herein should be reassessed and modified as appropriate.

During the modelling process we found the substrate layer biased towards rocky substrate (i.e., experts noted that rock was often reported in areas of known soft substrate). To a lesser degree, there are also predictions of soft substrate in known areas of rock. To manage this uncertainty, we assigned marginal suitability to unsuitable substrate categories (0.05 to 0.25 instead of 0) so that potentially suitable (but mischaracterised) habitat would not be designated as completely unsuitable.

Northern Abalone

Haliotis kamtschatkana

Based on the ecological knowledge outlined below, we included substrate, depth, salinity, tidal current speed, and exposure in the HSI model for Northern Abalone. Species distribution model predictions for Northern Abalone cannot be displayed due to its Species at Risk status. However, the relative influence of predictors for all three model types (HSI, GLM, BRT; Figure 4.3) shows that exposure and depth had the greatest influence on the HSI model prediction.

Northern Abalone generally occur from the low intertidal to around 100 m depth, with most occurring at depths of less than 10 m in BC waters (Sloan and Breen 1988). We assigned suitability as increasing from 0 to 1 between -6 m depth (high water line) and -3 m (mid-intertidal) to allow for a transition between unsuitable and fully suitable habitat. Suitability remained at 1 until 10 m, where we assigned an accelerated decrease in suitability to capture the fact that most of the population occurs above 10 m depth. We assigned 0 suitability after 100 m, the reported limit of the Northern Abalone depth distribution.

Northern Abalone are typically found on firm substrates such as rocks, boulders, and bedrock (DFO 2007), and optimum habitat consists of various combinations of ledges, cutbacks, depressions in stones, and boulder piles (Emmett and Jamieson 1988). Juvenile Northern Abalone (10 to 70 mm shell length) are found under and on exposed areas of rocks (DFO 2007) and are usually associated with crustose algae (Sloan and Breen 1988). Adults (> 70 mm shell length) are found on exposed rock surfaces (DFO 2007) and often colonize kelp beds (Williams 1989). We assigned a full suitability of 1 for rock, 0.5 for mixed, and 0.1 for sand and mud substrates.

Northern Abalone require full salinity (>30 ppt) seawater (COSEWIC 2009). We assigned salinity as unsuitable at values < 28 PSU (equivalent to ppt), then a linear increase in suitability between 28 and 30 PSU to allow for a transition between unsuitable and fully suitable salinities, with a full suitability of 1 thereafter.

When exposed to temperatures between 2 °C and 24 °C in the laboratory, Northern Abalone experienced no mortality or evidence of respiratory stress (Paul and Paul 1998). The range of average summer bottom temperatures within our study area fall entirely within the reported suitability range. Thus we did not include temperature as an explanatory variable.

Northern Abalone occur in a wide range of habitats from fairly sheltered bays to exposed coastlines (COSEWIC 2009), and are usually found in areas of moderate to high sea water exchange, such as in exposed or semi-exposed coastlines (DFO 2007). Growth is more rapid in moderately exposed than highly exposed areas, as strong wave action and water currents reduce the opportunities for abalone to catch and feed on drift algae (Sloan and Breen 1988). Studies in Barkley Sound and the Queen Charlotte Islands have found that translocating Northern Abalone from exposed to more sheltered habitat increases growth rates (Breen 1986; Emmett and Jamieson 1988), while several studies in BC have determined that Northern Abalone density increases with relative wave exposure, but mean shell length decreases (Breen and Adkins 1979; Lessard and Campbell 2007; Tomascik and Holmes 2003). Given that growth rates increase in more sheltered areas, the lower densities may be a result of historical fishers and poachers targeting areas that are easier to reach, and not decreased habitat suitability, however we were unable to find evidence to support this. We assumed that exposure was at least somewhat suitable over the entire range of values in the study area (0.005 to 84.9 *100 km) and assigned an optimum suitability of 1 for moderate to high exposure values between 30 to 70 (100s of km). We assigned an increasing relationship between 0 and 30, and a decreasing relationship between 70 and 84.9, to capture the fact that the ability to feed on drift kelp is likely

reduced in both low and high wave action. It should, however, be noted that the relationship at lower exposures is ambiguous.

While the influence of tidal currents on Northern Abalone is not well studied, Lessard and Campbell (2007) report that the species requires good water exchange, with tidal current or wave action present. We assumed that opportunities to capture and feed on drift algae would increase with higher current speeds, and assigned an increasing linear relationship with tidal current speed across the full range found in the study area. It is likely that drift algae capture would be reduced at higher current speeds, however the range of values within the study area for mean tidal current speed has a maximum value of 0.81 m·s-1 (1.57 knots), which is still relatively low and we assumed that feeding is not inhibited at this speed.

Given the preference for habitat with some structural complexity, especially for juveniles that shelter under rocks, we attempted to include a predictor for habitat complexity, first creating a suitability index for rugosity, and then one for slope. The resulting indices were highly restrictive, producing HSI models with very little suitable habitat. We presumed that the resolution of the predictor data were not high enough to capture the processes acting on Northern Abalone distribution. Therefore, we did not include habitat complexity in the model.

Pacific Geoduck

Panopea generosa

Based on the ecological knowledge outlined below, we included substrate, depth, salinity, tidal current speed, and exposure in the HSI envelope model for Pacific Geoduck (Figure A.2). The relative influence of predictors for all three model types (HSI, GLM, BRT) shows that substrate, depth, and exposure had the greatest influence on the HSI model prediction (Figure A.3).

We based the preferred depth distribution on Campbell et al. (1998) and dive density survey data collected within the study area. Campbell et al. (1998) report geoduck densities increase from 1 m depth to between 20 and 24 m depth, and then decrease to 110 m. This corresponds to data from the study area, which shows a peak in density at between 10 and 17 m, although data are limited at depths below 20 m (Bureau[1]). However geoduck beds are also found in intertidal areas in suitable substrate and conditions. Thus, we constructed an optimum depth preference with suitability increasing linearly from 0 to 1 between -2 and 0 m depth, full suitability until 25 m, and then a linear decrease in suitability to 0 at 110 m (this gives a value of 0.7 at 50 m depth, the limit of the study area).

As infaunal organisms, geoducks are primarily associated with soft bottom habitats, may occur on mixed substrate, and are absent on hard bottom (Goodwin and Pease 1989). We assigned a suitability of 1 to sand and mud substrates, 0.5 for mixed, and 0.25 for rock. Though geoduck are absent on hard bottom, the assigned value of 0.25 for rock accounts for an over-prediction of rock in the substrate layer, where some areas known to be soft substrate are predicted as rock.

Goodwin (1973) identified optimum bottom salinity and temperature ranges for larval survival. We used these values to define an optimum suitability curve, however the observed average summer bottom temperature in the study area was almost entirely within the optimal values thus this predictor was not included in the model.

High wind wave energy can result in poor substrate stability, reducing habitat suitability for geoducks. However, geoducks do seem to prefer some water movement as they are not found in highly sheltered areas, such as the heads of inlets (Bureau[1]). We therefore excluded very low and moderate to high exposure areas by assigning a linear increase in suitability from 0 to 1 between exposure values of 0 and 1 (100s of km), an optimum suitability of 1 until 25, then a linear decrease to 0 suitability at 50. We also ensured some water movement due to tides, excluding areas with a maximum tidal current speed < 3 $cm \cdot s^{-1}$, and then increasing habitat suitability to a maximum at 5 $cm \cdot s^{-1}$. The exposure and tidal values were developed in consultation with D. Bureau based on the position of known geoduck beds.

[1] Bureau, D. Research Biologist, Shellfish, Aquatic Resources, Research and Assessment Division (ARRAD), Science, Fisheries and Oceans Canada. March 2015 and December 2018, personal communication.

Red Sea Urchin

Mesocentrotus franciscanus

Based on the ecological knowledge outlined below, we included substrate, depth (conditioned on exposure), and salinity in the HSI envelope model for Red Sea Urchin (Figure A.5). The relative influence of predictors for all three model types (HSI, GLM, BRT) shows that depth had the greatest influence on the HSI model prediction (Figure A.6).

Urchins inhabit primarily rocky substrate, though can occur on mixed or soft substrate in low energy areas. We assumed a full suitability of 1 on rock substrate, 0.5 on mixed, 0.1 on sand, and 0.05 on mud.

Red Sea urchins are typically distributed from the intertidal to 50 m depth (DFO 2018), though individuals can be found to 284 m (Galloway[2]). Dan Leus[3] emphasized the importance of energy interacting with depth. This led to the separation of depth preference based on energy regime. To create the depth-exposure interaction, we assigned a low exposure depth relationship in areas where exposure values were 20 (100s of km) or less and a high exposure depth relationship in areas where exposure values were greater than 20 (100s of km). These relationships were combined into one predictor for depth, conditioned on exposure. For the low exposure depth relationship, we assigned depth as suitable across the range of -6 to 284 m, with a linear increase from 0 to 1 between -6 and -4 m depth, an optimum suitability of 1 from -4 to 9 m depth, and then a linear decrease to a marginal suitability of 0.1 by 14 m depth. For the high exposure depth relationship, we assigned depth as suitable across the range of -3 to 284 m, with a linear increase from 0 to 1 between 0 and 2 m depth, an optimum suitability of 1 from 2 to 15 m depth, and then a linear decrease to a marginal suitability of 0.1 by 20 m depth. For the high exposure depth relationship, the upper values of the range were shifted down, as the urchins will move deeper to avoid being washed away in high exposure areas. We assumed that this effect would not be seen below 20 m depth. In both high and low exposure depth relationships, there is a linear increase in suitability from 0 to 1 at the top of the depth range to capture the transition from unsuitable to suitable depths. Similarly, there is a transition from fully to marginally suitable just below the optimum depth range in both relationships.

Salinity was assigned as unsuitable below 28 PSU, with a linear increase in suitability between 28 and 31 PSU, and full suitability thereafter.

These urchins prefer moderate to strong currents (DFO 2018), as there will be increased food availability in these conditions, but avoid high-exposure areas where water pressure could turn them over and expose them to predation (Leus[3]). We included exposure as a condition for depth, where high exposure causes urchins to move deeper to avoid high wave action. Tidal current speed was not included in the model, as the predictors were not on an appropriate scale to resolve the ecological processes occurring.

[2] Galloway, A. Assistant Professor, University of Oregon, Oregon Institute of Marine Biology. October 2018, personal communication.

[3] Leus, D. Urchin Biologist, Marine Invertebrate Section, Stock Assessment & Research Division (StAR), Science, Fisheries and Oceans Canada. 2015 and December 2018, personal communication.

Pterygophora Kelp

Pterygophora californica

Based on the ecological knowledge outlined below, we included substrate, depth, salinity, tidal current speed, and exposure in the HSI envelope model for Pterygophora Kelp (Figure A.8). The relative influence of predictors for all three model types (HSI, GLM, BRT) shows that depth had the greatest influence on the HSI model prediction (Figure A.9).

Pterygophora is abundant on rocks and cobble (Druehl and Clarkston 2016; Hawkes et al. 1979) and we assumed a full suitability of 1 on rock substrate, 0.5 on mixed, and 0.1 on both sand and mud.

This species is found subtidally from 2 to 20 m depth (Mondragon and Mondragon 2003). We assigned depth as suitable over the range of 0 m (low water line) to 20 m, with a linear increase in suitability from 0 to 1 between 0 and 2 m depth, full suitability of 1 between 2 and 10 m, and then a delayed decrease to 0 suitability at 20 m. The linear increase from 0 to 2 m was included to give a transition between fully suitable and unsuitable, or marginally suitable habitat.

During preliminary review, it was suggested to use salinity tolerance values from Northern Abalone, as they are found in similar habitat as Pterygophora Kelp. Northern Abalone prefer full salinity seawater (>30 PSU) (COSEWIC 2009). We assigned salinity as unsuitable below 28 PSU, with increasing suitability between 28 and 30 PSU, and full suitability thereafter. The increasing suitability between 28 and 30 PSU was included to create a transition from unsuitable to fully suitable habitat.

This species typically becomes established on rocks in wave exposed areas (Druehl and Clarkston 2016) and is adapted for surf-swept and high current habitat (Lamb and Hanby 2006). Exposure was assigned as suitable over the range of 0 to 84.9 (100s of km), with a linear increase in suitability from 0 to 1 between 0 and 0.3, full suitability of 1 between 0.3 and 30, and then a linear decrease to 0 suitability at 84.9 (100s of km).

Tidal current speed was assigned as suitable over the range of 0 to 0.81 $m \cdot s^{-1}$, with a normal distribution across this range, giving a full suitability of 1 at 0.4 $m \cdot s^{-1}$.

Eelgrass

Zostera spp.

Based on the ecological knowledge outlined below, we included substrate, depth, salinity, temperature, tidal current speed, and exposure in the HSI model for Eelgrass (Figure A.11). The relative influence of predictors for all three model types (HSI, GLM, BRT) shows that depth and substrate had the greatest influence on the HSI model prediction (Figure A.12).

Eelgrass occurs from the mid-intertidal down to 30 m depth (Phillips and Lewis 1983). In consultation with experts, depth was assigned as suitable over the range of -6 m (high water line) to 30 m, with a linear increase in suitability from 0 to 1 between -6 and -3.3 m depth, full suitability of 1 between -3.3 and 10 m, and a linear decrease to a marginal suitability of 0.1 at 20 m, and then 0 suitability after 30 m depth. The linear increase and decrease in suitability from -6 to 3.3 m and 10 to 20 m (respectively) were included to give transitions between fully suitable and unsuitable, or marginally suitable habitat.

This species is found rooted in sandy or muddy sediments (Vandermeulen 2005) with an optimum habitat of mixed sand and mud (Phillips and Lewis 1983). For substrate, we assumed a full suitability of 1 on sand substrate, 0.75 on mud, 0.5 on mixed, and 0.25 on rock.

Eelgrass can tolerate a large range of temperatures from -6 to 40.5 °C, with an optimum temperature range of 10 to 20 °C (Phillips and Lewis 1983). Eelgrass can also tolerate a large range of salinities from freshwater to 42 PSU, with an optimum salinity range of 10 to 30 PSU (Phillips and Lewis 1983). We used the ranges and optima reported in the literature for both temperature and salinity. A full suitability of 1 was assigned across the optimum range, with a linear decrease in suitability down to 0 at the edges of the full range for each predictor.

Optimum habitat has light wave action (Phillips and Lewis 1983), and Eelgrass is common in the protected waters of bays and estuaries (Mondragon and Mondragon 2003). For exposure suitability, we assigned a decreasing linear relationship from 0 to 60 (100s of km), with a suitability of 0 thereafter. This relationship designated high exposure areas as unsuitable, as the optimum habitat for Eelgrass has light wave action.

Tidal current speed was assigned as suitable over the range of 0 to 0.81 $m \cdot s^{-1}$, with a normal distribution across this range, giving a full suitability of 1 at 0.4 $m \cdot s^{-1}$. This relationship ensures that both stagnant and high-current habitat were assigned marginal suitability.

Dungeness Crab

Metacarcinus magister

Based on the ecological knowledge outlined below, we included substrate, depth, salinity, and temperature in the HSI model for Dungeness Crab (Figure A.14). The relative influence of predictors for all three model types (HSI, GLM, BRT) shows that depth had the greatest influence on the HSI model prediction (Figure A.15), but influence of the other predictors was also quite high. This occurred because much of the study area had an overlap of full suitability for all predictors.

Dungeness crab occur from the low intertidal to 230 m depth (DFO 2017). A study by Stone and O'Clair (2001) from an Alaskan estuary reports variable depth preferences by gender and season ranging from less than 10 m to over 70 m, we thus assume a suitability of 1 across the full depth range for Dungeness (-2 to 230 m depth).

Dungeness crab typically inhabit substrates comprised of sand, mud, or silt (DFO 2017), although they have been observed on hard substrate (J. Lessard, personal observation). Given these preferences, we assign a suitability of 1 to soft bottom, 0.5 to mixed bottom, and 0.1 to hard bottom.

Despite having a large portion of their habitat in estuarine areas exposed to frequent episodes of low salinity, Dungeness crab are considered weak osmoregulators (Engelhardt and Dehnel 1973). They show behavioural and physiological signs of stress at salinities below 24 ppt (Curtis and McGaw 2012; McGaw et al. 1999) and are unable to tolerate prolonged exposure below 12 ppt (Cleaver 1949). Considering field data where Dungeness were rarely observed in salinities below 10 ppt (Curtis and McGaw 2008), we assign suitability increasing from 0 to 1 between 10 and 24 PSU (equivalent to ppt) and a full suitability of 1 thereafter.

In terms of water temperature, growth and survival are limited above 15°C due to increased energy expenditure associated with respiration (Kondzela and Shirley 1993). The upper lethal temperature limit for Dungeness crab ranges from 20 to 25 °C, depending on the season (Pauley et al. 1989). In a lab experiment, where Curtis and McGaw (2012) challenged Dungeness crab with starvation and temperature and salinity stresses, crabs tolerated temperatures as high as 15°C. We assign an optimum suitability of 1 for temperatures up to 15 °C, with suitability decreasing to 0 by 25 °C.

Dungeness crab are active predators of shellfish, typically clams, and may also catch fish. They have a strong preference for fresh food but will scavenge dead prey (Curtis 2019[4]). Dungeness crab predation in estuaries is a balancing act between physiological challenges and the availability of clam prey that typically occur in shallower, less saline waters. This implies that on a more regional scale, estuaries of the kind preferred by clams would be important Dungeness habitat. For crab, their estuarine habitat is a function of salinity, temperature, and exposure, as accessibility of the clam prey depends on stratified waters that allow the formation of a saline wedge during a rising tide (Curtis 20194). There is currently no exposure dataset available at the correct spatial resolution for the shelf study area, so we attempted to use tidal current speed to highlight areas of low water movement where this saline wedge may form. We found that the 100 m by 100 m resolution used in this study area was too coarse to resolve fine scale processes such as salinity wedge formation. Thus, tidal current speed was not included in the model.

[4] Curtis, D. Research Biologist, Shellfish Section, Stock Assessment and Research Division, Fisheries and Oceans Canada. January 2019, personal communication.

Quillback Rockfish

Sebastes maliger

Based on the ecological knowledge outlined below, we included substrate, depth, salinity, and temperature in the HSI model for Quillback Rockfish (Figure A.17). The relative influence of predictors for all three model types (HSI, GLM, BRT) shows that substrate had the greatest influence on the HSI model prediction (Figure A.18), but influence of the other predictors was also quite high. This occurred because much of the study area had an overlap of full suitability for all predictors.

During review of the preliminary model, one suggestion was to create separate models for juveniles and adults. Given time constraints, however, we have modeled juveniles and adults together and will create separate models at a later date.

Quillback Rockfish are distributed from the Kenai Peninsula in the Gulf of Alaska to the Anacapa Passage in southern California (Love et al. 2002), and depth is generally reported to be the most important determinant of distribution (National Marine Fisheries Service 2014). Across its range, this species has been reported subtidally to 274 m depth (Love et al. 2002). In BC, Quillback observations range from 9 to 186 m depth (98% of all survey data in Ground Fish Bio database – top and bottom 1% excluded) and newly settled juveniles have likely been observed in Eelgrass beds and other intertidal areas (Copper, Quillback, and Brown Rockfish juveniles are indistinguishable; Haggarty[5]). Considering all of these reported ranges, we assigned a full depth suitability of 1 from 0 to 168 m depth, with suitability decreasing to 0 by 274 m.

Quillback aggregate over hard, complex substrates with some vertical relief, such as broken rock, rock reefs, ridges, and crevices (Richards 1986) and juveniles may also be seen over cobble and soft substrates, such as in eelgrass beds (Richards and Hand 1987)**.** We thus assign a full suitability of 1 for rock and mixed substrates, 0.5 suitability for sand, and a marginal suitability of 0.1 for mud.

Within BC, temperatures from Quillback observations in bottom trawl surveys range from 5.78 to 23 °C and salinities from Quillback observations in bottom trawl (2013 to 2018) and longline surveys (2018) range from 28.2 to 33.9 PSU (Keppel[6]). Temperature and salinity ranges reported for Quillback and Yelloweye Rockfish in southeastern Alaska (Johnson et al. 2003) near the northern end of their distribution fall within the temperature ranges observed in BC, and we were unable to find reported temperature or salinity ranges near the southern portion of their distribution. Given a lack of information on preferences and optima, we assign a full suitability of 1 across the temperature mean and salinity mean ranges observed for BC.

Quillback prefer complex habitats such as fractured bedrock and boulder complexes (e.g., National Marine Fisheries Service 2014; Yamanaka et al. 2006). To capture this habitat complexity, we first created a suitability index for rugosity, and then one for slope. The resulting indices were highly restrictive, producing HSI models with very little suitable habitat, as data were not on a fine enough scale to capture the processes acting on Quillback distribution. We therefore chose not to include rugosity and slope indices in the model.

[5] Haggarty, D. Inshore Rockfish and Lingcod Program Head, Groundfish Section, Stock Assessment and Research Division, Fisheries and Oceans Canada. December 2018, personal communication.

[6]Keppel, E. Biologist, Groundfish Section, Stock Assessment and Research Division, Fisheries and Oceans Canada. December 2018, personal communication.

Yelloweye Rockfish

Sebastes ruberrimus

Based on the ecological knowledge outlined below, we included substrate, depth, salinity, and temperature in the HSI model for Yelloweye Rockfish (Figure A.20). The relative influence of predictors for all three model types (HSI, GLM, BRT) shows that substrate had the greatest influence on the HSI model prediction (Figure A.21), but influence of the other predictors was also quite high. This occurred because much of the study area had an overlap of full suitability for all predictors.

During review of the preliminary model, one suggestion was to create separate models for juveniles and adults. Given time constraints, however, we have modeled juveniles and adults together and will create separate models at a later date.

Yelloweye are generally distributed from Umnak and Unalaska Islands in the Aleutian Islands to Ensenada in northern Baja California and have been reported from 15 to 549 m depth (Love et al. 2002). In BC waters, inshore Yelloweye observations range from 31 to 250 m depth (98% of all survey data in Ground Fish Bio database – top and bottom 1% excluded). Considering these reported ranges, we assigned depth suitability as increasing from 0 to 1 between 15 and 31 m to give a transition from unsuitable to fully suitable depth. Suitability remains at 1 from 31 to 250 m, then decreases to 0.1 by 300 m. Suitability remains marginal at 0.1 until 549 m depth, after which suitability is 0. The decrease between 250 and 300 m allows for a transition between fully and marginally suitable depth, and fits with feedback from preliminary review that depth suitability likely decreases faster than a linear decrease between 250 and 549 m.

Yelloweye prefer rocky habitat, but may occupy benthic substrates such as sand and mud (National Marine Fisheries Service 2014). We thus assign a full suitability of 1 for rock and mixed substrates, 0.25 suitability for sand (Haggarty), and a marginal suitability of 0.1 for mud.

Within BC, temperatures observed from Yelloweye observations in bottom trawl surveys range from 4.56 to 16.8 °C and salinities from Yelloweye observations in bottom trawl (2013 to 2018) and longline surveys (2018) range from 28.3 to 34 PSU (Keppel[6]). Temperature and salinity ranges reported Yelloweye in southeastern Alaska (Johnson et al. 2003) near the northern end of their distribution fall within the temperature ranges observed in BC, and we were unable to find reported temperature or salinity ranges near the southern portion of their distribution. Given a lack of information on preferences and optima, we assign a full suitability of 1 across the temperature and salinity ranges observed for BC.

Yelloweye prefer complex habitats such as fractured bedrock and boulder complexes (e.g., National Marine Fisheries Service 2014; Yamanaka et al. 2006). To capture this habitat complexity, we first created a suitability index for rugosity, and then one for slope. The resulting indices were highly restrictive, producing HSI models with very little suitable habitat, as data were not on a fine enough scale to capture the processes acting on Yelloweye distribution. We therefore chose not to include rugosity and slope indices in the model.

Ochre Sea Star

Pisaster ochraceus

Based on the ecological knowledge outlined below, we included substrate and depth in the HSI model for the Ochre Sea Star (Figure A.22). The relative influence of predictors shows that depth had the greatest influence on the model prediction (Figure A.23).

Ochre sea stars occur from the mid-intertidal down to 87 m depth (Harbo 2011), with an optimum range from -3 m (mid-intertidal) to 2 m (Harley[7]). We assigned depth as suitable across the range of -5 to 87 m, with a linear increase from 0 to 1 between -5 and -3 m depth, an optimum suitability of 1 from -3 to 2 m depth, and then a linear decrease to a marginal suitability of 0.1 by 10 m depth. There is a linear increase in suitability from 0 to 1 at the top of the depth range to capture the transition from unsuitable to suitable depths. Similarly, there is a transition from fully to marginally suitable just below the optimum depth range.

These sea stars inhabit rocky coastline and are ecologically significant predators in a broad range of environments, from sheltered lagoons to the most wave-exposed shorelines (Dayton 1971; Levin and Paine 1974). For substrate, we assigned a full suitability of 1 for rock, moderate suitability of 0.5 for mixed, and marginal suitabilities of 0.1 for sand and 0.05 for mud substrates. Given that ochre sea stars can tolerate a wide range of wave and tidal energies, tidal current speed and exposure were not included in the model.

Ochre Sea Stars can also be found in a wide range of salinities and temperatures. In a study investigating the effect of salinity changes on low- and high-salinity populations, Held and Harley (2009) used populations found naturally occurring at 20 PSU (Vancouver, BC) and 30 PSU (Bamfield, BC). The authors found that the low salinity populations were able to survive for extended periods at 15 PSU, the lowest salinity tested, and that each of the populations had the highest feeding rates at their *in situ* salinity. In a study investigating the effects of temperature, Gooding et al. (2009) found that growth rates increased linearly over temperatures from 5 to 21 °C. However, Sanford (2002) also found greater growth per gram of mussel tissue consumed at 9 °C than at 12 °C, suggesting that reduced consumption under colder conditions was balanced by reduced metabolic costs. The lethal limit for body temperature is 35.8 °C (Pincebourde et al. 2008). Given that the reported ranges are within those found in the study area, temperature and salinity were not included in the model. Even though growth rates are reported to increase linearly over temperatures from 5 to 16 °C, it is unclear what the optimum range is, and we assume all of these temperatures are suitable for survival.

[7] Harley, C. Professor, UBC Department of Zoology and Institute for the Oceans and Fisheries. April 2019, personal communication.

Blue Mussel complex

Mytilus edulis, *M. trossulus*, and *M. galloprovincialis*

Based on the ecological knowledge outlined below, we included substrate, depth, and tidal current speed in the HSI model for the Blue Mussel complex (Figure A.24). The relative influence of predictors shows that depth had the greatest influence on the model prediction (Figure A.25).

In the Pacific, there are three species or sub-species considered to be part of the *Mytilus edulis* complex, *M. edulis*, *M. galloprovincialis*, and *M. trossulus*. These species cannot be distinguished based on shell features alone, and are able to hybridize (Harbo 2011). We modeled these species together.

Harbo (2011) reports that these mussels occur from the intertidal to 5 m depth and Williams (1989) reports a maximum depth of 45 m, with dense colonization between -3.7 (high mid-intertidal) and -1.5 m (low intertidal) depth. Depth was assigned as suitable across the range from -6 to 45 m, with a linear increase in suitability from 0 to 1 between -6 and -5 m depth, full suitability of 1 from -5 to -1.5 m (low intertidal) depth, and then a linear decrease to a marginal suitability of 0.1 by 5 m depth. The increase in suitability at the top of the depth range and the decrease in suitability between -1.5 and 5 m depth were included to capture transitions between fully and marginally suitable habitat. During preliminary review of the predictors, reviewers suggested that the optimal depth range for blue mussels would be where the small acorn barnacle, *Chthamalus dalli,* occurs (Norgard and Bigg[8]). Klinkenberg (2017) reports that *C. dalli* occurs high in the intertidal zone, and Morris et al. (1980) report high to mid-intertidal occurrence. To capture this range, we extended the optimum depth range for blue mussel reported by Williams (1989; -3.7 to -1.5 m depth) to include the upper intertidal (up to -5 m depth).

Blue mussels form dense aggregations on hard surfaces (Harbo 2011) and can be found on rocks, pier pilings, floats, gravel, compact mud, and hard-shelled organisms (Carlton 2007; Morris et al. 1980; Williams 1989). For substrate, we assigned a full suitability of 1 for rock, moderate suitability of 0.5 for mixed, and a marginal suitability of 0.1 for sand and mud substrates.

The mussels can tolerate a wide range of salinities and temperatures, from 15 to 40 PSU (Williams 1989) and 10 to 25 °C (Brenko and Calabrese 1969), respectively. Salinity and temperature were not included in the model because values in the study area were within the reported ranges.

Blue Mussels can also tolerate a wide range of wave and tidal energies, occurring from sheltered estuaries to exposed coastline (Williams 1989). However, reviewers indicated that some locations within the study area are too stagnant for survival (Norgard and Bigg[8]). We used tidal current speed to exclude stagnant habitat, and assigned increasing suitability from 0 to 1 between 0.03 and 0.05 $m \cdot s^{-1}$, with a full suitability of 1 thereafter. Exposure was not included in the model, as blue mussels are found in a wide range of wave energies and we were able to exclude the lowest energy regimes using the tidal current speed predictor relationship.

[8] Norgard, T., and Bigg, M. Marine Spatial Ecology and Analysis (MSEA) section, Science, Fisheries and Oceans Canada. December 2018, personal communication.

Littleneck Clam

Leukoma staminea

Based on the ecological knowledge outlined below, we included substrate, depth, tidal current speed, and exposure in the HSI model for Littleneck Clam (Figure A.26). The relative influence of predictors shows that depth had the greatest influence on the model prediction (Figure A.27).

Littleneck Clam inhabits areas with stable sand, packed mud, or gravel clay mixtures, and is sometimes found in gravelly sediments among rocks (Williams 1989). We assigned a full suitability of 1 for sand and mud substrates, moderate suitability of 0.5 for mixed substrate, and a marginal suitability 0.1 for rock.

Harbo (2011) reports that this clam inhabits the mid intertidal down to 10.5 m depth, and Williams (1989) reports a depth range of slightly above mid-intertidal to 12 m. For depth, we used a delayed decreasing relationship, assuming that these clams occur preferentially at the shallow end of their depth range, where nutrients and dissolved oxygen are available in higher concentrations. We assigned suitability as increasing from 0 to 1 between -6 m (high water line) and -3 m (mid-intertidal) depth, and then a delayed decrease, reaching 0 suitability at 12 m depth.

Adult Littleneck Clams can survive in temperatures ranging from 0 to 25 °C (Chew and Ma 1987) and salinities from 20 to 35 PSU (Glude 1978; Quayle and Bourne 1972). The range of temperature mean and salinity mean values in our study are were entirely within the reported temperature and salinity ranges, thus no temperature or salinity suitability index was required in the model.

The Littleneck Clam is common on protected beaches in bays and estuaries where exposure is low and wave and current energy are low to moderate (Williams 1990). For exposure values, we created a threshold relationship, with a full suitability of 1 across low exposure values up to 30 (100s of km) then a linear decrease to 0 suitability at the moderate exposure value of 50. This excluded high exposure areas where soft substrates would likely be washed away. Similarly, we kept tidal current speed values low to moderate, with a suitability of 1 up to 0.4 $m \cdot s^{-1}$ then a linear decrease to 0 suitability at 0.81 $m \cdot s^{-1}$.

Orange Sea Pen

Ptilosarcus gurneyi

Based on the ecological knowledge outlined below, we included substrate, depth, and tidal current speed in the HSI model for the Orange Sea Pen (Figure A.28). The relative influence of predictors shows that depth had the greatest influence on the model prediction (Figure A.29).

The Orange Sea Pen is a sessile colonial anthozoan found in dense aggregations in soft sediment habitat (Lamb and Hanby 2006). Given this preference, we assign a suitability of 1 for sand and mud substrates and 0.5 for mixed substrate. In the substrate predictor layers for the shelf study area there are several areas where rock is predicted in areas known to be primarily sand or mud. To account for this, we assign a suitability of 0.25 for rock substrate so that these areas are not predicted as completely unsuitable.

Orange Sea Pens are reported to occur from the intertidal to 100 m depth or more by Harbo (2011) and subtidally to 135 m depth by Lamb and Hanby (2006). Combining these reports, and in the absence of information on depth preferences, we assume a suitability of 1 across their full reported depth range from the low intertidal (-1 m) to 135 m depth.

Orange Sea Pens are passive suspension feeders and rely on ambient currents to deliver food to a network of polyps on leaf-like branches on the upper part of the organism. In a series of laboratory and field experiments near Friday Harbour, Washington, Best (1988) examined the relationship between ambient (tidal) velocity and feeding efficiency in the Orange Sea Pen. They estimated feeding efficiencies increased to maxima between 8 and 20 $cm \cdot s^{-1}$, depending on the size of the individual, outside of which efficiency decreased. Based on these estimates, we assign a parabolic relationship for velocity with an optimum suitability of 1 occurring at 14 $cm \cdot s^{-1}$.

We do not include salinity or temperature predictors, as there is a lack of information on thresholds and preferences for these two environmental variables for this species.

APPENDIX B REFERENCES

Best, B.A. 1988. Passive suspension feeding in a sea pen: Effects of ambient flow on volume flow rate and filtering efficiency. Biological Bulletin **175**: 332-342.

Breen, P., and Adkins, B. 1979. A survey of abalone populations on the east coast of the Queen Charlotte Islands, August 1978. Fish. Mar. Serv. Manuscr. Rep **1490**: 125.

Breen, P.A. 1986. Management of the British Columbia fishery for Northern Abalone (*Haliotis kamtschatkana*). Can. Spec. Pub. Fish. Aquat. Sci. **92**: 300-312.

Brenko, M.H., and Calabrese, A. 1969. The combined effects of salinity and temperature on larvae of the mussel *Mytilus edulis*. Mar. Biol. **4**(3): 224-226.

Campbell, A., Harbo, R.M., and Hand, C.M. 1998. Harvesting and distribution of Pacific geoduck clams, *Panopea abrupta*, in British Columbia. *In* Proceedings of the North Pacific Symposium on Invertebrate Stock Assessment and Management. *Edited by* G.S. Jamieson and A. Campbell. NRC Research Press, Ottawa, ON. pp. 349-358.

Carlton, J.T. 2007. The Light and Smith Manual: Intertidal Invertebrates from Central California to Oregon. University of California Press, Berkeley, California.

Chew, K.K., and Ma, A.P. 1987. Species profiles: Life histories and environmental requirements of coastal fishes and invertebrates (Pacific Northwest) - common littleneck clam. U.S. Fish and Wildlife Service. 22 pp.

Cleaver, F.C. 1949. Preliminary results of the coastal crab (*Cancer magister*) investigation. Department of Fisheries, Washington State. 47–92 pp.

COSEWIC. 2009. COSEWIC assessment and update status report on the Northern Abalone, *Haliotis kamtschatkana*, in Canada. Committee on the Status of Endangered Wildlife in Canada, Ottawa, ON. vii + 48 pp.

Curtis, D.L., and McGaw, I.J. 2008. A year in the life of a Dungeness crab: methodology for determining microhabitat conditions experienced by large decapod crustaceans in estuaries. J. Zool. **274**: 375–385.

Curtis, D.L., and McGaw, I.J. 2012. Salinity and thermal preference of Dungeness crabs in the lab and in the field: Effects of food availability and starvation. Journal of Experimental Marine Biology and Ecology **413**: 113-120.

Dayton, P.K. 1971. Competition, disturbance, and community organization: the provision and subsequent utilization of space in a rocky intertidal community. Ecological Monographs **41**: 351-389.

DFO. 2007. Recovery Strategy for the Northern Abalone (*Haliotis kamtschatkana*) in Canada. *Species at Risk Act* Recovery Strategy Series. Fisheries and Oceans Canada, Vancouver. *vi* + 31 pp.

DFO. 2017. Pacific Region Integrated Fisheries Management Plan: Crab by Trap - January 1, 2017 to March 31, 2018.

DFO. 2018. 2018/2019 Red Sea Urchin Integrated Fisheries Management Plan.

Druehl, L.D., and Clarkston, B. 2016. Pacific Seaweeds: A guide to the common seaweeds of the West Coast. Harbour Publishing, Madeira Park, BC.

Emmett, B., and Jamieson, G.S. 1988. An experimental transplant of northern abalone, *Haliotis kamtschatkana*, in Barkley Sound, British Columbia. Fishery Bulletin, U.S. **87**: 95-104.

Engelhardt, F.R., and Dehnel, P.A. 1973. Ionic regulation in Pacific edible crab, *Cancer magister* (Dana). Canadian Journal of Zoology **51**: 735–743.

Glude, J.B. 1978. The clam genera *Mercenaria, Saxidomus, Protothaca, Tapes, Mya, Panope,* and *Spisula*: a Literature review and analysis of the use of thermal effluent in the culture of clams. Aquaculture Consultant Report. 74 pp.

Gooding, R.A., Harley, C.D.G., and Tang, E. 2009. Elevated water temperature and carbon dioxide concentration increase the growth of a keystone echinoderm. PNAS **106**(23): 9316-9321.

Goodwin, C.L. 1973. Effects of salinity and temperature on embryos of the geoduck clam (*Panopea generosa*). Proceedings of the National Shellfish Association. pp. 93-95.

Goodwin, C.L., and Pease, B. 1989. Species profiles: life histories and environmental requirements of coastal fishes and invertebrates (Pacific Northwest) - Pacific geoduck clam. U.S. Fish and Wildlife Service 11.120.

Harbo, R.M. 2011. Whelks to Whales: Coastal Marine Life of the Pacific Northwest. Harbour Publishing, Madiera Park, B.C.

Harley, C. Professor, UBC Department of Zoology and Institute for the Oceans and Fisheries. April 2019, personal communication.

Hawkes, M.W., Tanner, C.E., and Lebednik, P.A. 1979. The benthic marine algae of northern British Columbia. Syesis **11**: 81-115.

Held, M.B.E., and Harley, D.G.H. 2009. Responses to low salinity by the sea star *Pisaster ochraceus* from high- and low-salinity populations. Invertebr. Biol. **128**(4): 381–390.

Johnson, S.W., Murphy, M.L., and Csepp, D.J. 2003. Distribution, habitat, and behavior of rockfishes, *Sebastes* spp., in nearshore waters of southeastern Alaska: observations from a remotely operated vehicle. Environ. Biol. Fishes **66**: 259–270.

Klinkenberg, B. 2017. E-Fauna BC: Electronic Atlas of the Fauna of British Columbia. Available from efauna.bc.ca.

Kondzela, C.M., and Shirley, T.C. 1993. Survival, feeding, and growth of juvenile Dungeness crabs from southeastern Alaska reared at different temperatures. J. Crust. Biol. **13**: 25–35.

Lamb, A., and Hanby, B. 2006. Marine Life of the Pacific Northwest: A Photographic Encyclopedia of Invertebrates, Seaweeds and Selected Fishes. Harbour Publishing, Madiera Park, B.C.

Lessard, J., and Campbell, A. 2007. Describing northern abalone, *Haliotis kamtschatkana*, habitat: focusing rebuilding efforts in British Columbia, Canada. J. Shellfish Res. **26**(3): 677-686.

Levin, S.A., and Paine, R.T. 1974. Disturbance, patch formation, and community structure. PNAS **71**: 2744-2747.

Love, M.S., Yoklavich, M., and Thorsteinson, L. 2002. The Rockfishes of the Northeast Pacific. University of California Press, Berkeley, California.

McGaw, I.J., Reiber, C.L., and Guadagnoli, J.A. 1999. Behavioral physiology of four crab species in low salinity. Biological Bulletin **196**: 163–176.

Mondragon, J., and Mondragon, J. 2003. Seaweeds of the Pacific Coast: Common Marine Algae from Alaska to Baja California. Sea Challengers, Monterey, CA. pp. 97.

Morris, R.H., Abbott, D.L., and Haderlie, E.C. 1980. Intertidal invertebrates of California. Stanford University Press, Stanford, CA.

National Marine Fisheries Service. 2014. Designation of Critical Habitat for the Distinct Population Segments of Yelloweye Rockfish, Canary Rockfish, and Bocaccio. NMFS, Protected Resources Division, West Coast Region, Seattle, WA. 77 pp.

Paul, A.J., and Paul, J.M. 1998. Respiration rate and thermal tolerances of pinto abalone (*Haliotis kamtschatkana*). J. Shellfish Res. **17**: 743-745.

Pauley, G.B., Armstrong, D.A., Van Citter, C., and Thomas, G.L. 1989. Species profiles: life histories and environmental requirements of coastal fishes and invertebrates (Pacific Southwest)—Dungeness crab. U.S. Fish and Wildlife Service. Biological Report 11.121.

Phillips, R.C., and Lewis, R.L. 1983. Influence of environmental gradients on leaf widths and transplant success in North American seagrasses. Mar. Technol. Soc. J. **17**: 59-68.

Pincebourde, S., Sanford, E., and Helmuth, B. 2008. Body temperature during low tide alters the feeding performance of a top intertidal predator. Limnol. Oceanogr. **23**: 1562-1573.

Quayle, D.B., and Bourne, N. 1972. The clam fishery of British Columbia. Fisheries Research Board of Canada Bulletin **179**: 70.

Richards, L.J. 1986. Depth and habitat distributions of three species of rockfish (*Sebastes*) in British Columbia: observations from the submersible PISCES IV. Environ. Biol. Fishes **17**(1): 13-21.

Richards, L.J., and Hand, C.M. 1987. 1987 research catch and effort data on nearshore reef-fishes in British Columbia statistical areas 12 and 13. Can. Man. Rep. Fish. Aquat. Sci. **1958**:59.

Sanford, E. 2002. The feeding, growth, and energetics of two rocky intertidal predators (*Pisaster ochraceus* and *Nucella canaliculata*) under water temperatures simulating episodic upwelling. J. Exp. Mar. Biol. Ecol. **273**: 199–218.

Sloan, N.A., and Breen, P.A. 1988. Northern Abalone, *Haliotis kamtschatkana*, in British Columbia: Fisheries and Synopsis of Life History Information. Can. Spec. Pub. Fish. Aquat. Sci. **103**.

Stone, R. P., and C. E. O'Clair (2001), Seasonal movements and distribution of Dungeness crabs, Cancer magister, in a glacial southeastern Alaska estuary, Mar. Ecol. Prog. Ser., **214:** 167-176.

Tomascik, T., and Holmes, H. 2003. Distribution and abundance of *Haliotis kamtschatkana* in relation to habitat, competitors and predators in the Broken Group Islands, Pacific Rim National Park Reserve of Canada. J. Shellfish Res. **22**: 831-838.

Vandermeulen, H. 2005. Assessing marine habitat sensitivity: a case study with eelgrass (*Zostera marina* L.) and kelps (Laminaria, Macrocystis). DFO Can. Sci. Advis. Sec. Res. Doc.. 2005/032 57 pp.

Williams, G.L. 1989. Coastal/estuarine fish habitat description & assessment manual Part I: Species/Habitat Outlines. Williams & Associates Ltd., Coquitlam, BC. 139 pp.

Williams, G.L. 1990. Coastal/estuarine fish habitat description & assessment manual Part II: Habitat Description Procedures. Williams & Associates Ltd., Coquitlam, BC. 61 pp.

Yamanaka, K.L., Lacko, L.C., Miller-Saunders, K., Grandin, C., Lochead, J.K., Martin, J.C., Olsen, N., and Wallace , S.S. 2006. A review of quillback rockfish, *Sebastes maliger*, along the Pacific coast of Canada: biology, distribution and abundance trends. DFO Can. Sci. Advis. Sec. Res. Doc. 2006/077. 68 pp.