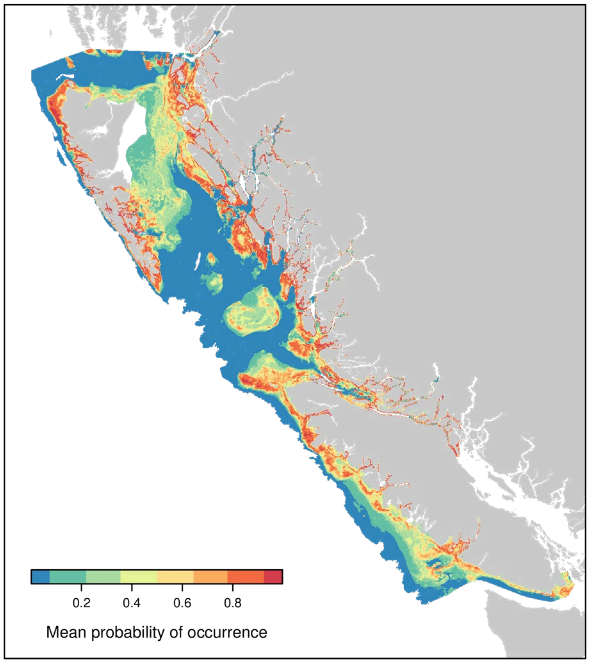
Fisheries and Oceans Canada | Pêches et Océans Canada

Ecosystems and Oceans Science | Sciences des écosystèmes et des océans

DEVELOPMENT OF A SPECIES DISTRIBUTION MODELLING FRAMEWORK AND ITS APPLICATION TO TWELVE SPECIES ON CANADA'S PACIFIC COAST



An example predicted species distribution map.

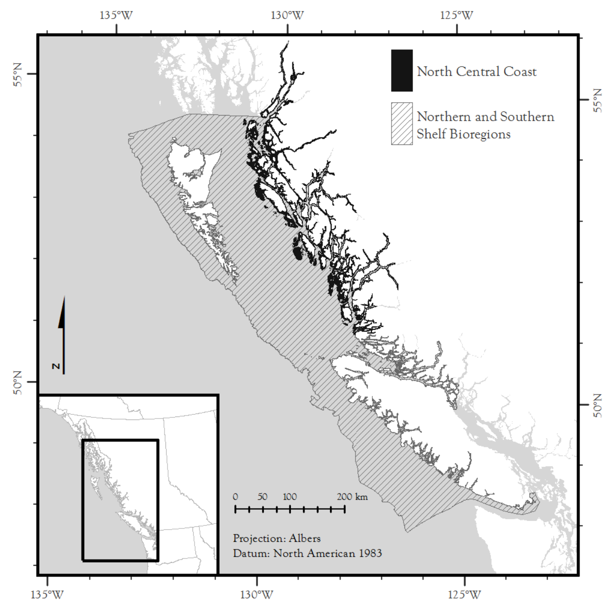


Figure 1. The two study areas used in the application of the framework. The nearshore (20 m resolution) North Central Coast and shelf (100 m resolution) Northern and Southern Shelf Bioregions.

Context:

Species distribution models (SDMs) can be used to predict the distribution of species by relating observations of species occurrence to environmental data. Understanding species' distributions can inform a variety of management activities including marine spatial planning and emergency pollution response. SDMs can also identify gaps in ecological knowledge, helping to target future survey and research efforts. However, SDMs can be misapplied if species or environmental data are not appropriately screened and prepared, and proper model validation is not applied (Elith and Leathwick 2009; Hawkins et al. 2003; Roberts et al. 2016).

Fisheries and Oceans Canada (DFO) Science Branch proposed this peer review to provide a comprehensive and standardized approach to building SDMs based on best practices to ensure consistent quality and rigour in their use and application. The assessment and advice arising from this Canadian Science Advisory Secretariat (CSAS) Regional Peer Review (RPR) will be used to develop SDMs, and integrate them, where appropriate, into science and policy decisions related to the management and conservation of marine species.

This Science Advisory Report is from the June 11-12, 2019 regional peer review on Habitat Suitability Modelling Best Practices for Canada's Pacific Ocean. Additional publications from this meeting will be posted on the [Fisheries and Oceans Canada (DFO) Science Advisory Schedule](#) as they become available.

January 2020 (Errata: October 2020)

Canada

SUMMARY

- Species distribution models (SDMs) are relevant to several national objectives related to marine spatial planning, vulnerability assessments, emergency response, and stock assessment. Benefits to Fisheries and Oceans Canada (DFO) of effectively applying SDM methods include explicit consideration of ecological and management contexts, consistent preparation of available data, and application of appropriate analytical methods to construct and evaluate models.
- DFO Pacific Science Branch has proposed a framework that includes a set of guidelines for the development of consistent, interpretable, and defensible SDMs. The framework can be implemented with purpose-built scripts (https://gitlab.com/dfo-msea/sdm) written in the R statistical programming language (R Core Team 2018).The framework is intended to support practitioners by following current best practices and providing guidance on key aspects of data preparation, model fitting and evaluation, uncertainty estimation, and interpretation of results.
- An application of the framework is illustrated by applying three modelling methods of increasing complexity to twelve benthic species, for two study areas on Canada's Pacific coast (Figure 1). The resulting predictions are evaluated using standardized performance metrics, and diagnostic plots and maps, including the relative importance and marginal effects of contributing predictors. The application demonstrates the importance of a consistent data model, building multiple models including a knowledge-based model, and uncertainty estimation.
- SDMs for the twelve species are intended to support oil-spill response planning and preparedness. The models generally fit the data well as indicated by the performance metrics. The appropriateness of these model predictions for uses beyond oil spill response will need to be assessed by the user, based on clearly identified model objectives defined by their particular application.
- Sources of uncertainty arise from methodological choices, and the availability and quality of data. Uncertainty in the model predictions is assessed with four different methods that reduce or measure uncertainty spatially across a study area. Additional sources of uncertainty are considered by following best practices for data selection where practical.
- The framework is a scientifically defensible, transparent, and reproducible method for producing SDMs. It can be used to produce additional species models, or adapted for other modelling extents.
- Several recommendations for the successful development and application of SDMs arose from the framework.

BACKGROUND

Overview of species distribution modelling

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, SDMs have become a common approach to estimating distributions of valued species. However, there are several challenges to building SDMs. These include the increasing accessibility of 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.

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 (Figure 2). Since the management of a species or its habitat is inseparable from its ecology, the ecological model describes both the ecological and management contexts. Its definition begins with the best practice of clearly articulating the objectives of the model, and includes an explicit description of what is known about the species and its ecology. This informs the necessary extents and resolution of the model, and construction of the data model. The data model is comprised of species observations and environmental predictors presumed to 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 relevant interactions between them. The statistical model is used to relate observations of a species to predictor data, and to predict species' distribution across the study area. Statistical methods range broadly in complexity. The primary criteria for selecting statistical methods are whether it can meet objectives and how well it makes use of the available data.

Data selection considerations

All data come with biases and all models make assumptions. For example, observations of species may occur in only a portion of the species' range, while assumptions about a species' movement will determine whether model results are relevant year-round, or in a particular season. Best practice for SDMs requires the consideration of such biases and assumptions when selecting appropriate species and environmental data for modelling.

Data on observations of species come in different forms and are collected in many ways. Targeted or non-random survey designs can lead to predictions of patterns in the data collection rather than the species' distribution, especially if the sampling is spatially or environmentally structured (Araújo et al. 2019). Observations that span as much of the spatial and temporal extents of the species' distribution as possible are preferred. Reduced spatial, temporal, and taxonomic precision of occurrence data will limit modelling options. Combining observations from different data sources to increase sample size and extents should be considered, but care is required to avoid obscuring sampling biases as this may lead to misleading results and increased model uncertainty.

Predictors describing the marine environment are available from a variety of sources from remote sensing technologies (e.g., satellite imagery, drones, acoustics) and ocean models (including elevation, bottom type, and ocean dynamics). Spatial variables, such as distance to important physical or biotic features (e.g., shoreline) should also be considered. Additionally, fishing effort can be an important predictor of species distribution in impacted ecosystems (e.g., Foster et al. 2015; Tien et al. 2017). The resolution of marine predictors can span orders of magnitude, ranging from the kilometre scale for oceanographic model predictors to the sub-metre scale for acoustically-derived topographic predictors. Best practice requires predictor resolution to match the scale at which it is presumed to influence the species' distribution (e.g., Wiens 1989). The inevitable mismatch in resolution between observations and predictors (and often among predictors) means that some predictor variables are likely to represent an ecological process well, while others will serve as proxies for predictors operating at different resolutions. How important this mismatch is will depend on the model objectives, primarily whether the model is required to be transferable (i.e., provide reliable predictions in places or at times beyond the study area), or only explain the observed pattern within the study area.

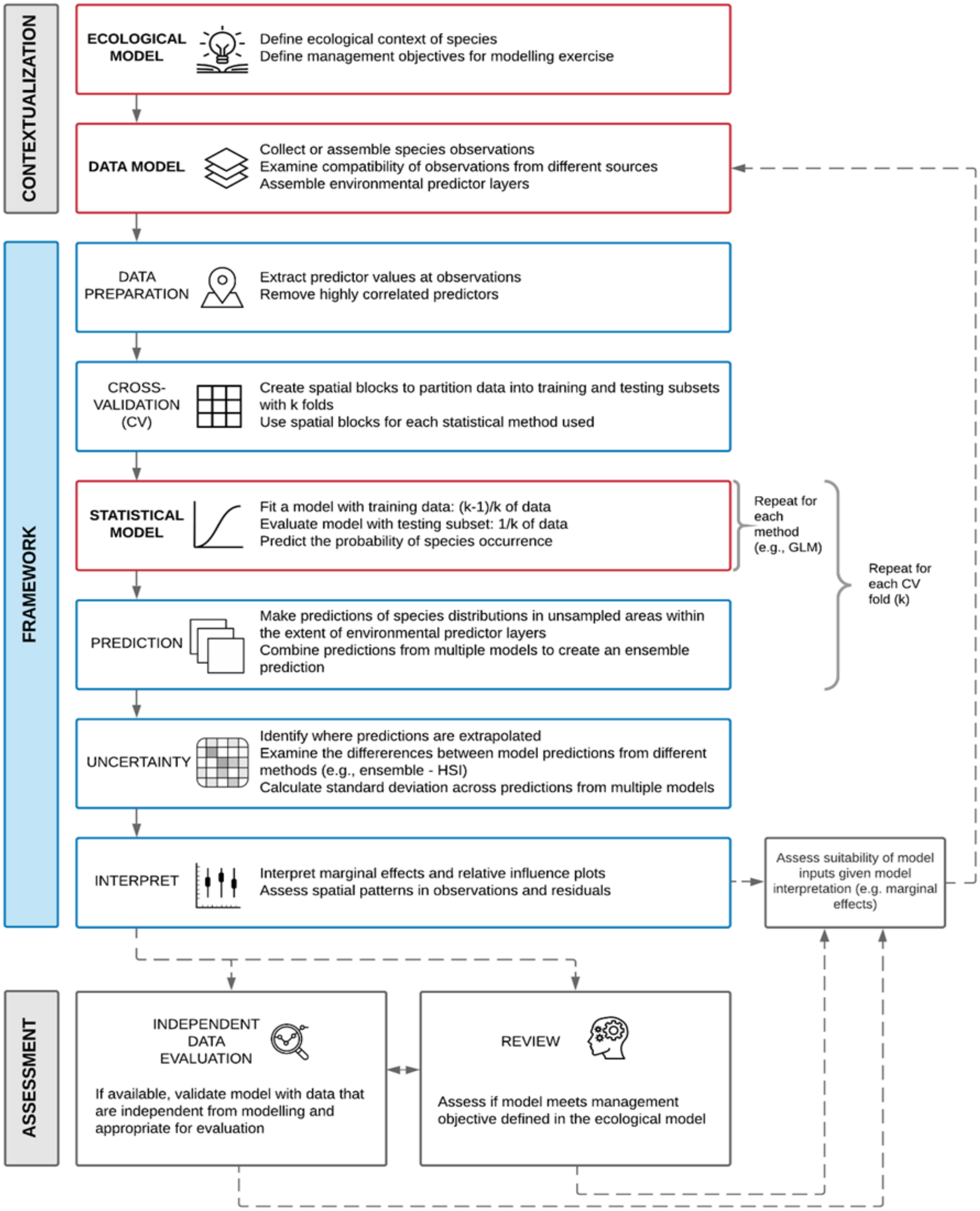


Figure 2. Overview of the complete modelling process from developing the model context (Contextualization) to the assessment of the model predictions. The Framework developed here for species distribution modelling includes a series of prescribed steps that automate best practices. Generalized linear 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.

ASSESSMENT

Overview of the SDM framework

The SDM framework was developed with model building best practices in mind, and refined through experience obtained during its application. To implement the framework, purpose-built scripts (https://gitlab.com/dfo-msea/sdm) were written in the R statistical programming language (R Core Team 2018). The framework is comprised of six steps (Figure 2): data preparation, cross-validation, model fitting, model evaluation, prediction, and interpretation. Models are evaluated using a spatial block cross-validation approach to increase the independence of observations used to fit (training data) and evaluate (testing data) the models. Model fitting is done repeatedly with a variety of methods to create multiple models that are combined into an ensemble prediction. Ensemble predictions are more robust (Oppel et al. 2011) and allow insights into model uncertainty. Uncertainty is reduced by limiting extrapolation. It is evaluated by measuring the variation among model predictions that make up the ensemble, comparing a knowledge-based model prediction to a data-driven model prediction, and by mapping model residuals. Models are interpreted by examining the relative influence of predictor variables in the model and the marginal effects of each predictor on the observations.

The framework focuses on data-driven SDMs (e.g., generalized linear models). It also considers the utility of knowledge-based envelope models, which provide a way of bounding species' distributions using hypothesized environmental constraints. They are used in the framework 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.

Data and methods used for the application

Twelve benthic species were selected for modelling to illustrate the application of the framework, and guide emergency oil spill response planning as part of the Regional Response Plan. The framework was applied to the shelf waters of Canada's Pacific Coast, excluding the Strait of Georgia, at two resolutions: 100 m for the Northern and Southern Shelf Bioregions, and 20 m for the nearshore, North Central Coast region (Figure 1). Environmental predictor layers used in the models (Table 1) included bathymetry and related derivatives, physical, chemical and biological oceanography, fetch, and substrate. The predictors were a mix of interpolated measured variables, predictions from oceanographic models, and topographic models of varying complexity.

The twelve benthic species selected are conservation priorities (Gale et al. 2019) or highly vulnerable to oil (Hannah et al. 2017).They were selected 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. Occurrence data for the twelve species (Table 2) were sourced from targeted systematic surveys. For some species, more than one survey was used to source occurrence observations. This was needed when sample sizes were low (e.g., < 100 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). To reduce sampling bias, observations from different surveys were only combined if the sampling gear types were the same. Four of the twelve species selected (Ochre Sea Star, Blue Mussel, Littleneck Clam and Orange Sea Pen) did not have adequate occurrence data for data-driven modelling within our study areas.

The application used three model fitting methods that span the range of model complexity: habitat suitability index (HSI) models, generalized linear models (GLMs), and boosted regression trees (BRTs). HSIs are knowledge-based envelope models that depend on published information and expert consultation to bound species' distributions based on environmental conditions (Brooks 1997; US Fish and Wildlife Services [USFWS] 1981). GLMs and BRTs are statistical methods that relate occurrence data to environmental predictors (Elith and Graham 2009). Following the framework methods, predictions from data-driven models (e.g., GLM and BRT) were combined into an ensemble prediction using a performance-weighted mean.

The occurrence data used for model fitting and evaluation was divided into five spatially distinct datasets, or blocks, to perform cross validation. Five GLM, BRT and ensemble models were built and evaluated. The mean and standard deviation of a set of the evaluation metrics and prediction surfaces were calculated across the five models. Model performance was measured using the area under the receiver operator characteristic curve metric which ranges from 0 to 1. Scores less than 0.5 indicate models that are worse than random and scores of 1 indicate that the model perfectly predicts the occurrence data (Freeman and Moisen 2008; Merckx et al. 2011). For each species, all models were evaluated with the same data to facilitate model comparison. However, development of the HSI models differed from the data-driven models in that only one HSI model was built for each species. Additionally, only HSI models could be built for species deemed to be data deficient (Table 2).

Table 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).

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 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

Table 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). Source data surveys (ABL = Abalone; RSU = Red Sea Urchin; BHM = benthic habitat mapping; Cuke = Sea Cucumber; HBLL = hard bottom long line; IPHC = International Pacific Halibut Commission) represent stock assessments or monitoring led by DFO Pacific and industry partners. For this application, data were sourced from those surveys were presence-absence (P-A) or absence only (Abs). Study area indicates nearshore (N) or shelf (S); Obs/n indicates the prevalence of observations for each species.

Species	Study area	Sample size (n)	Obs/n (%)	Years	Source (P-A)	Source (Abs)	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[1]	3	2010-2017	ABL, Cuke, RSU, BHM[2]	-	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[3]	4	2010-2017	GDK, Cuke, RSU, BHM[4]	-	Points	Estimated location along transect
Dungeness Crab (*Metacarcinus magister*)	S	391[5]	49[6]	1982-2009[7]	Dungeness Crab survey[8]	-	Points	Start position of gear deployment [9]
Quillback Rockfish (*Sebastes maliger*)	S	4,937[10]	41[11]	2003-2018[12]	HBLL, IPHC	-	Lines	Start and end position of longline gear
Yelloweye Rockfish (*Sebastes ruberrimus*)	S	4,937[10]	51	2003-2018[12]	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	-	-	-	-	-	-

[1] Erratum: October 2020 - 6,486 now reads 6,607
[2] Erratum: October 2020 - ABL, Cuke, BHM now reads ABL, Cuke, RSU, BHM
[3] Erratum: October 2020 - 12,650 now reads 12,567
[4] Erratum: October 2020 - ABL, Cuke, BHM now reads GDK, Cuke, RSU, BHM
[5] Erratum: October 2020 - 6,702 now reads 391
[6] Erratum: October 2020 - 10 now reads 49
[7] Erratum: October 2020 - 1975-2018 now reads 1982-2009
[8] Erratum: October 2020 - Crab, Shrimp trawl, BHM now reads Dungeness Crab surveys
[9] Erratum: October 2020 - Start of transect (Crab, Shrimp trawl) or estimated position on transect (BHM) now reads Start position of gear deployment
[10] Erratum: October 2020 - 4,758 now reads 4,937
[11] Erratum: October 2020 - 40 now reads 41
[12] Erratum: October 2020 - 2003-2017 now reads 2003-2018

Results from the application

Maps of occurrence observations, model predictions and uncertainty measures, and the relative influence and marginal effects of the environmental predictor variables were generated for all species. Results for Quillback Rockfish are included as an example (Appendix A).

The best performing ensemble models were Quillback Rockfish and Abalone (Figure 3). Models of Geoduck, Kelp, Urchin and Eelgrass all performed moderately well. The poorest performing models were Dungeness Crab and Yelloweye Rockfish, and although their performance was relatively poor, model predictions still performed better than random.

Collectively, HSI models performed poorly with the majority of these models performing below the worst data-driven models. The poorest performing HSI model was Eelgrass which was no better than random. The differences in model performance are likely driven by differences in the ecological models and the distribution and quality of available occurrence data rather than simply the sample size of observations.

For species where HSI model performance was particularly poor, comparisons of marginal effects and relative influence plots (e.g., Figure A3) point to potentially misunderstood or improperly scaled drivers. The identification of species-specific problems with the ecological or data models demonstrates the diagnostic utility of the performance metrics generated by the framework. For example, unevenness in the relative influence of predictors across cross-validation models implies model misspecification, while differences in model predictions can indicate uncertainty in model structure and resolution. The use of multiple modelling methods thus provides a means to assess confidence in model structure and predictor appropriateness.

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 (see Figure A2). For the four data deficient species, the preliminary envelope models provide a baseline from which data-driven models of distribution can be developed once observational data are available.

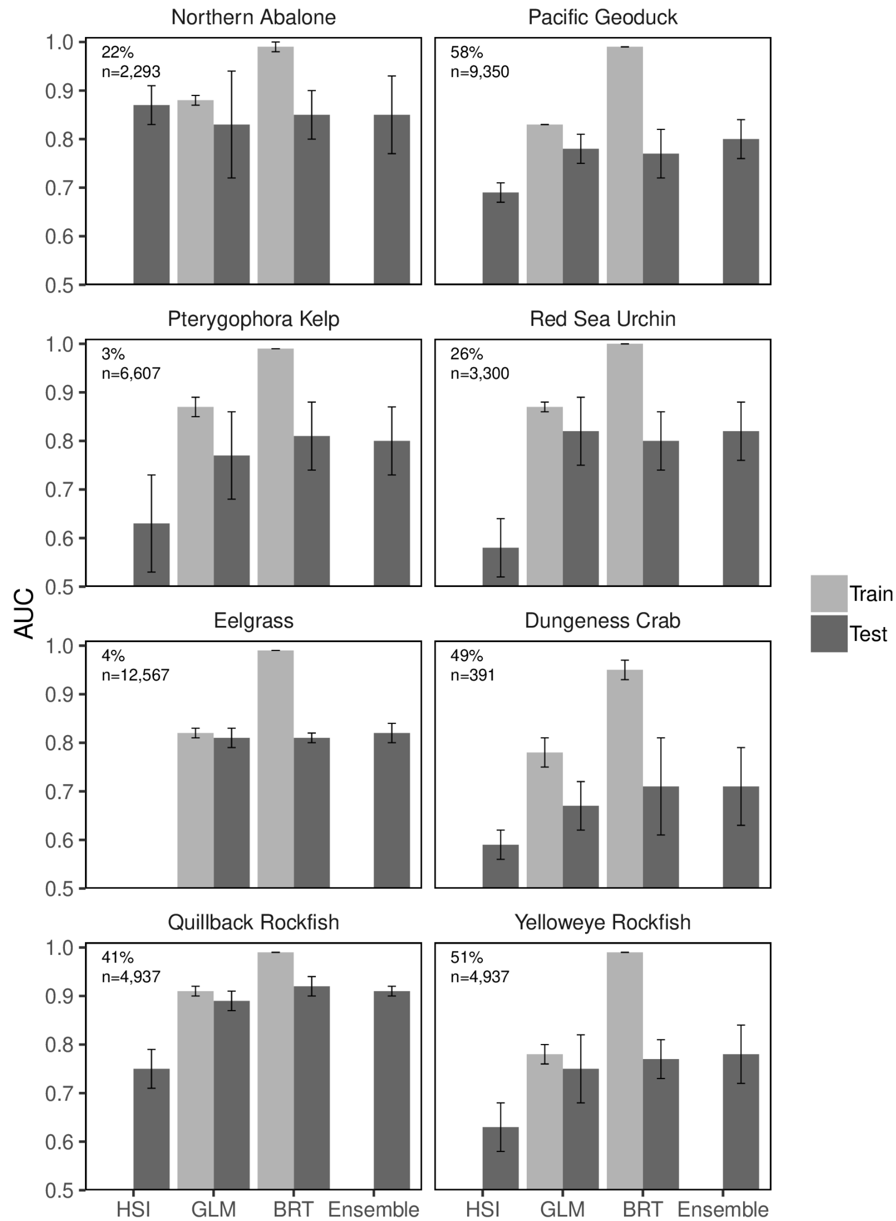


Figure 3. Performance of the eight distribution models for species with sufficient occurrence data. Mean area under the receiver operator characteristic curve (AUC) is based on five-fold spatial cross-block validation tests and are shown for each of the four model types: habitat suitability index (HSI), 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.

Sources of Uncertainty

- Uncertainty arises from the availability and quality of data. Sampling bias in the occurrence data is a main source of uncertainty and can be confounded if occurrence data are combined from multiple sources. This raises the question of when it is appropriate to combine data from different sources. Uncertainty is minimized by following best practice advice for data selection where practical.

- Interpolation within the study area is generally appropriate for models that perform well. If extrapolation to other times and places is deemed appropriate, then it should be done with caution.
- Estimates of confidence in model predictions as well as insights into aspects of data quality and sampling bias can be gained through the comparison of knowledge-based and data-driven model predictions.
- A multi-model ensemble approach facilitates estimates of prediction uncertainty by allowing a comparison between models built with different methods.
- The variation among predictions from models built using different subsets of the observations (using spatial block cross-validation) provides insight into spatial sampling bias and can highlight areas of lower certainty. Locations where the modelled relationships are less certain may be an indication of local non-stationarity.
- Uncertainty stemming from errors in the predictor layers, spatial precision of observations, missing predictors, or scale mismatches between the predictors and observations are considered during the data preparation phase prior to the framework but are not captured in any numeric uncertainty estimate. In addition, model parameter uncertainty is not captured in the uncertainty assessments.

CONCLUSIONS AND ADVICE

The presented Research Document along with the formal peer review meeting and this Science Advice present an appropriate review of background materials along with challenges and best practices for species distribution model development.

The framework is a scientifically defensible, transparent, and reproducible method for producing SDMs. It is suitable for producing additional species models, or adaptation to other study areas. The framework also facilitates updates when new occurrence data or improved predictors became available. This iterative approach to re-assessing model inputs for their suitability during the model interpretation phase and subsequently refining the models (see Figure 2) led to improved model performance in many cases.

A multi-model approach facilitates uncertainty estimates by allowing a comparison between different model types. In the application presented, 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 also helped inform reliability assessments for individual species models.

The framework has been applied to twelve species of variable data quality and quantity. The models generally fit the data well as indicated by the performance metrics. The twelve SDMs will be used to inform emergency response in the event of a marine oil spill incident. The appropriateness of these model predictions for uses beyond oil spill response will need to be assessed by the user, according to their particular application.

The 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 such species, additional observations obtained through a well-designed sampling program would improve the model predictions. For species with high-quality observational data, modelling species density would be a valuable next step to provide additional information for oil spill response and other management needs.

Several recommendations for the successful development and application of SDMs arose from the framework.

- Clearly identify model objectives as they inform the development of the data and statistical models and are central to interpreting model results.
- Take care in selecting species and environmental data, as modelling with biased data can lead to poor or misleading results.
- Incorporate local ecological knowledge wherever possible; such knowledge is especially valuable in data poor situations.
- Build knowledge-based envelope models, as they clarify the current ecological understanding, help identify important environmental predictors, and provide a means of estimating model uncertainty.
- Build ensemble predictions from multiple models as the ensemble approach produces robust predictions and provides a means for estimating model uncertainty.
- Use independently collected data for additional model validation where practical.

The Research Document (Nephin et al., 2019) that accompanies this process contains much more detailed and explicit technical discussions of considerations in implementing species distribution modelling and is a valuable resource for practitioners.

LIST OF MEETING PARTICIPANTS

Last Name	First Name	Affiliation
Anderson	Sean	DFO Science, Stock Assessment
Beazley	Lindsay	DFO Science, Maritimes
Benoy	Nicholas	DFO Ecosystem Management, Oceans
Campbell	Jill	DFO Science
Candy	John	DFO Science, Centre for Science Advice
Chiang	Eric	DFO Oceans
Christensen	Lisa	DFO Science, Centre for Science Advice
Curtis	Janelle	DFO Science, Ecosystems Sciences
Davies	Sarah	DFO Science, Ecosystems Sciences
Dudas	Sarah	DFO Science, Ecosystems Sciences
English	Philina	DFO Science, Stock Assessment
Ferguson	Kiyomi	DFO Science
Fields	Cole	DFO Science, Ecosystems Sciences
Finney	Jessica	DFO Science, Ecosystems Sciences
Gale	Katie	DFO Science, Ecosystems Sciences
Gomez	Catalina	DFO Science, Maritimes
Goulet	Pierre	DFO Science, Newfoundland Region
Gregr	Edward	University of British Columbia

Last Name	First Name	Affiliation
Gullage	Lauren	DFO Science, Newfoundland Region
Herborg	Matthias	DFO Science, Ecosystems Sciences
Hubley	Brad	DFO Science, Maritimes Region
Kenchington	Ellen	DFO Science, Maritimes
Knudby	Anders	University of Ottawa
Lessard	Joanne	DFO Science, Ecosystems Sciences
Leus	Dan	DFO Science, Ecosystems Sciences
Murillo-Perez	Javier	DFO Science, Maritimes
Nephin	Jessica	DFO Science, Ecosystems Sciences
Novaczek	Emilie	DFO Science, Newfoundland Region
O	Miriam	DFO Science, Ecosystems Sciences
Pretty	Christina	DFO Science, Newfoundland
Poehlke	Travis	DFO Ecosystem Management, Oceans
Robinson	Cliff	DFO Science, Ecosystems Sciences
Rooper	Chris	DFO Science, Stock Assessment
Rubidge	Emily	DFO Science, Ecosystems Sciences
Sameoto	Jessica	DFO Science, Maritimes
Schut	Steven	DFO Science, Stock Assessment
St. Germain	Candice	DFO Science, Oceans Science Division
Warren	Margaret	DFO Science, Newfoundland Region
Wells	Nadine	DFO Science, Newfoundland
Yakgujaanas	Jaasaljuus	Council of Haida Nations

SOURCES OF INFORMATION

This Science Advisory Report is from the June 11-12, 2019 regional peer review on Habitat Suitability Modelling Best Practices for Canada's Pacific Ocean. Additional publications from this meeting will be posted on the Fisheries and Oceans Canada (DFO) Science Advisory Schedule as they become available.

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.

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

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).

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 + 38p.

Elith, J., and Leathwick, J.R. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual review of ecology, evolution, and systematics 40: 677-697.

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.

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.

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.

Gale, K.S.P., Frid, A., Lee, L., McCarthy, J., Robb, C., Rubidge, E., Steele, J., and Curtis, J.M.R. 2018. 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.

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.

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

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.

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.

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

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.

Nephin, J., Gregr, E.J., St. Germain, C., Fields, C., and Finney, J.L. 2019. 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. 2019/nnn. In Press.

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.

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

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.

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.

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

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

APPENDIX A. RESULTS FROM QUILLBACK ROCKFISH SPECIES DISTRIBUTION MODEL

This appendix presents an example of the modelling results from the application of the SDM framework on Canada's Pacific Coast with Quillback Rockfish as an example species. For the modelling results for all other species in the application see Nephin et al. (2019). Figure A1 represents the distribution of the presence and absence observations used for modelling, Figure A2 shows the predictions for the distribution of Quillback Rockfish within the study area and its related uncertainty, and Figure A3 displays the relative influence and marginal effects of the environmental predictor variables.

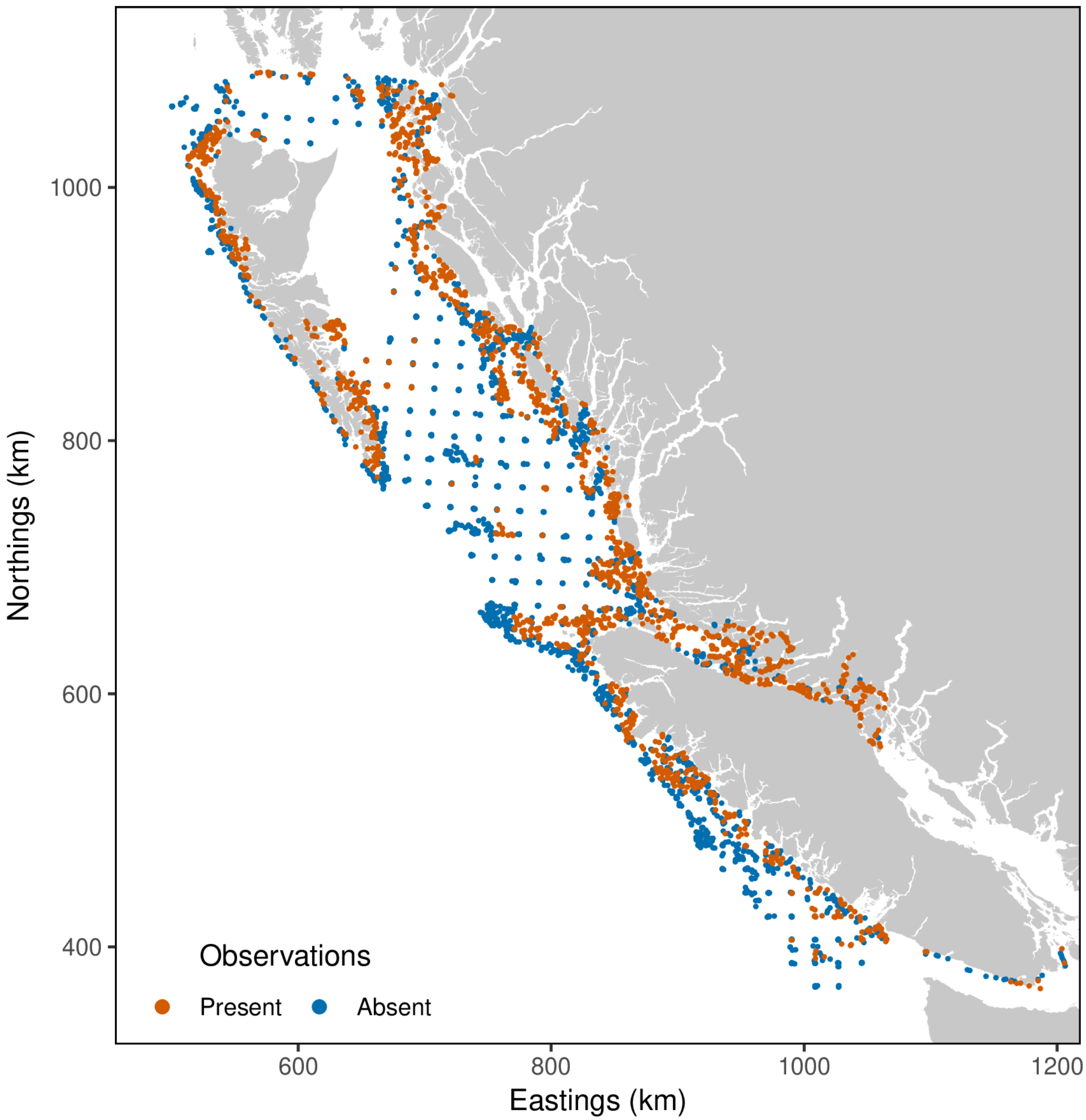


Figure A.1 Quillback Rockfish presence and absence observations within the shelf study area.

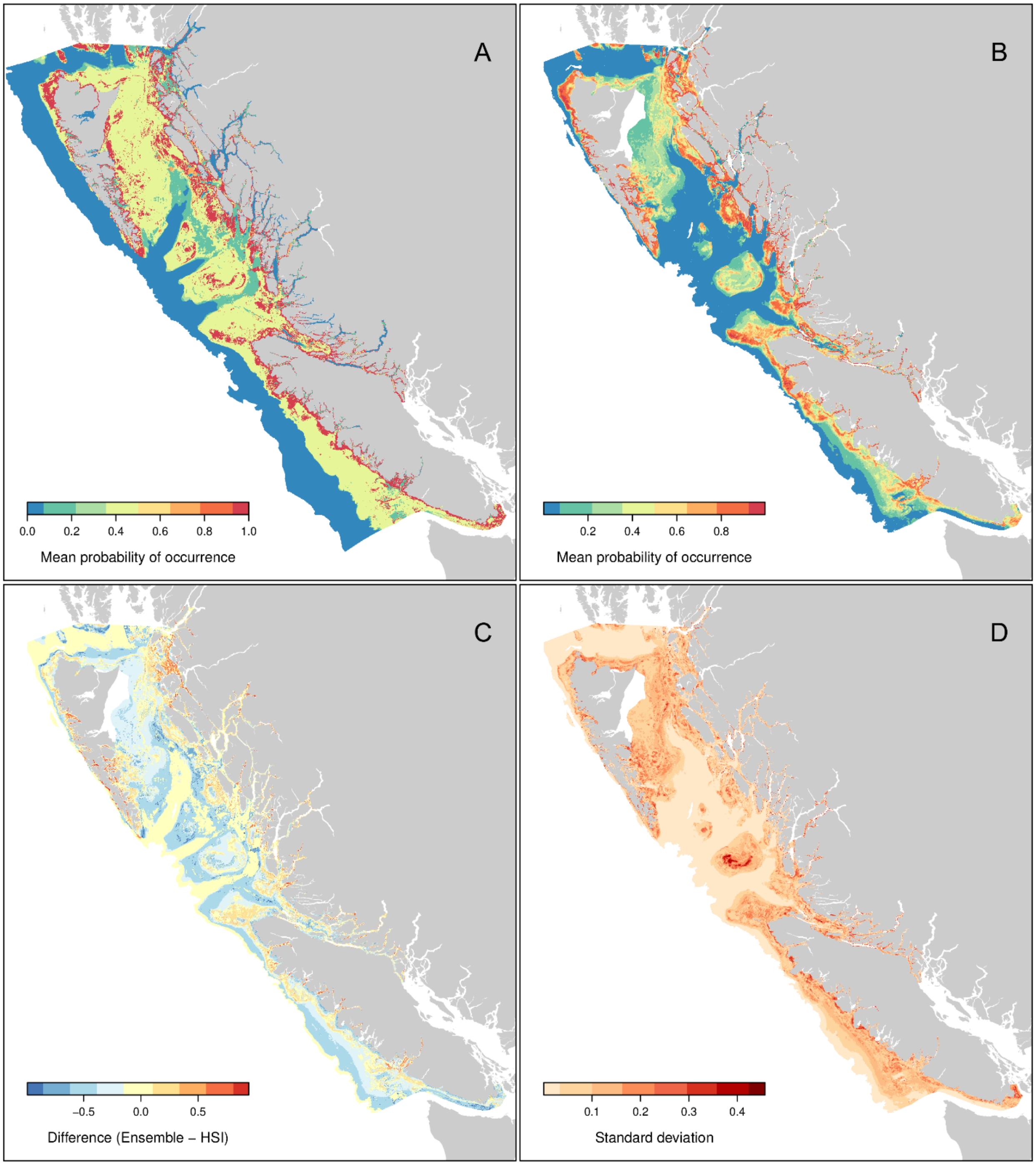


Figure A.2. Predictions of Quillback Rockfish distribution and the related uncertainty. Probability of occurrence predictions from A) the habitat suitability index model (HSI) and B) the ensemble model based on generalized linear and boosted regression tree models. Model uncertainty is represented by C) the difference between the HSI and the ensemble model predictions and D) the standard deviation across multiple ensemble model predictions.

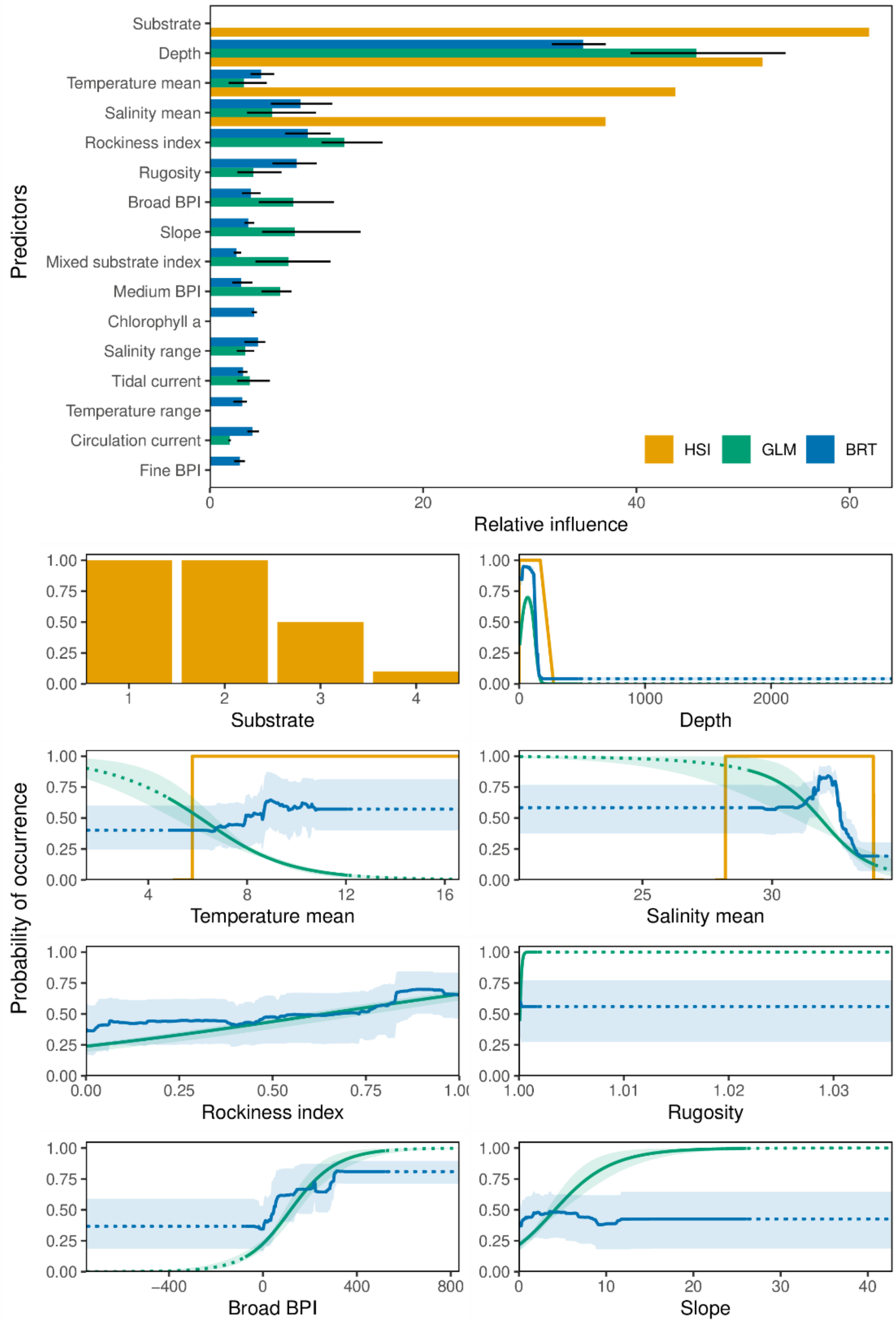


Figure A.3. Relative influence of predictors (top) and marginal effects (bottom, multi-panel) for the eight most influential environmental predictors from each of the HSI, GLM and BRT Quillback Rockfish models. For GLM and BRT models, the bars in the relative influence plots represent the mean and the 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.

THIS REPORT IS AVAILABLE FROM THE:

Centre for Science Advice (CSA)
Pacific Region
Fisheries and Oceans Canada
3190 Hammond Bay Road
Nanaimo, BC V9T 6N7

Telephone: (250) 756-7208
E-Mail: csap@dfo-mpo.gc.ca
Internet address: www.dfo-mpo.gc.ca/csas-sccs/

ISSN 1919-5087

Correct Citation for this Publication:

DFO. 2020. Development of a Species Distribution Modelling Framework and its Application to Twelve Species on Canada’s Pacific Coast. DFO Can. Sci. Advis. Sec. Sci. Advis. Rep. 2020/004. (Errata: October 2020)

Aussi disponible en français :

MPO. 2020. Élaboration d’un cadre de modélisation de la répartition des espèces et son application à douze espèces de la côte canadienne du Pacifique. Secr. can. de consult. sci. du MPO, Avis sci. 2020/004. (Errata : Octobre 2020)